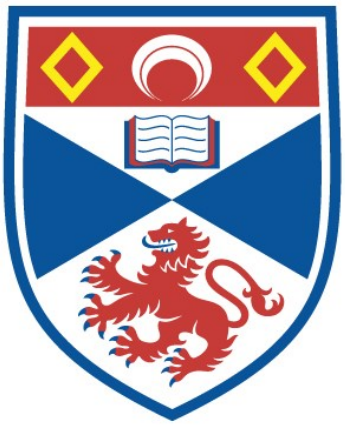
THE EVOLUTION OF GALAXY STRUCTURES FROM A BAYESIAN PERSPECTIVE

Joshua J. Argyle

A Thesis Submitted for the Degree of PhD
at the
University of St Andrews



2017

The Evolution of Galaxy Structures from a Bayesian perspective

Photometric decompositions of galaxies in the near and distant universe

by

Joshua J Argyle

University of
St Andrews

FOUNDED
1413

Submitted for the degree of Doctor of Philosophy in Astrophysics

May 2017

Declaration

I, Joshua Argyle, hereby certify that this thesis, which is approximately 50,000 words in length, has been written by me, that it is the record of work carried out by me and that it has not been submitted in any previous application for a higher degree.

Date Signature of candidate

I was admitted as a research student in September 2013 and as a candidate for the degree of PhD in September 2013; the higher study for which this is a record was carried out in the University of St Andrews between 2013 and 2017.

Date Signature of candidate

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of PhD in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date Signature of supervisor

Copyright Agreement

In submitting this thesis to the University of St Andrews we understand that we are giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. We also understand that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that my thesis will be electronically accessible for personal or research use unless exempt by award of an embargo as requested below, and that the library has the right to migrate my thesis into new electronic forms as required to ensure continued access to the thesis. We have obtained any third-party copyright permissions that may be required in order to allow such access and migration, or have requested the appropriate embargo below.

The following is an agreed request by candidate and supervisor regarding the electronic publication of this thesis: Access to Printed copy and electronic publication of thesis through the University of St Andrews.

Date Signature of candidate

Date Signature of supervisor

Abstract

Galaxy structures in the local Universe are the result of an evolution spanning billions of years. The diversity in morphologies observed is due to mechanisms that could either be from external interactions or internal processes. The hierarchical merging of two massive galaxies has long been thought to give rise to pressure-supported spherical structures, including elliptical galaxies and classical bulges. On the other hand, isolated galaxies may evolve at a much slower pace with the accretion of gas forming flattened rotationally-supported discs. Secular evolution could also result in newly formed internal structures, such as bars or discy-bulges. The first step in understanding the complex pathways that formed these monolithic beasts, we need to robustly measure their structures.

This thesis investigates how the structure of galaxies have evolved over the last seven Gyrs. In the first part I present a new Bayesian Markov Chain Monte Carlo (MCMC) two-dimensional (2D) photometric decompositions algorithm called `PHI`. The purpose of the algorithm is to decompose a galaxy's light profile into the various components that make-up the structure. By implementing a three level method, `PHI` is able to obtain a full understanding of the parameter space overcoming many of the major issues previous codes have struggled with.

The second part of the thesis describes the generation of synthetic galaxy images which are used to test the robustness of `PHI` and to also highlight the cosmological and instrumental effects that may bias the outcomes. We also present is a performance test for the Bayesian application to bulge-disc decompositions of galaxies using images from the Sloan Digital Sky Survey (SDSS), also included is the parameter estimation and model comparison method using the Bayesian Information Criteria.

In the third part of this thesis we show how the use of Hierarchical Bayesian Models (HBM) can be used to describe structural scaling relations in the local Universe. A constant piece-wise representation fully captures the underlying nature of the sample. This leads to the analysis of many structural scaling relations, one being the positive trend between the effective radius of the bulge and the Sérsic index.

Lastly, a study investigating the structural evolution of galaxies within the COSMOS field in the redshift range $0.5 < z < 1.25$ is presented. The flexible nature of the HBM allows for a detailed description of the build-up of galaxy structures in the local Universe.

Acknowledgements

The completion of this thesis would not have been possible without the encouragement, support and help of many people.

Firstly, I would like to thank my supervisors: Dr. Jairo Méndez-Abreu and Dr. Vivienne Wild. You both have supported me in every aspect of my academic life, and made my time here in St Andrews deeply rewarding. The combination of your relaxed nature, encouraging comments, and sometimes brutal honesty, without a doubt, has helped my future career prospects.

Particular thanks goes to my friends and colleagues in the astronomy department. A special thanks goes to my officemates, Inna, Maya, Gabi, with a hug and a kiss for Tim and sturdy handshake for David. There is never a dull moment when Alistair and Alasdair are around, as well as making it very difficult to say no to the occasional pub visit. A massive combined thanks to the 'super' official St Andrews Astronomy Squash Society, you know who you are. I can't let David get off that easy without a second thank you for being my (sometimes unwilling) house-maid, I mean mate. It has been a fantastic few years and I wish you all the best in finding someone else to finish your video games.

Most importantly, I would like thank my parents, grandparents, brother Callum, and the pups. Without the support, guidance and love from you all I honestly don't think there would have ever been a chance that I'll be submitting a PhD thesis. It is probably due to the collective of: mini-math problems, the toy telescope and the devout viewings of Star Wars that unwittingly set the ball rolling all them years ago. *May the force be with you*.

Finally, I am especially grateful to Maria for being exactly who you are. In the vastness of space and the immensity of time, it is my joy to share a spot with you, preferably in a handstand.

Contents

List of Figures

List of Tables

1 Introduction

1.1 Historical concepts of galaxies

Lost somewhere between immensity and eternity, is where Sagan (1981) placed us in the opening chapters of the supporting material to his award-winning television series, 'Cosmos: A Personal Voyage'. The Milky Way was believed to occupy more or less the whole Universe at the beginning of the twentieth century. The idea that extragalactic systems could exist was not established until the 1920's, with the pioneering efforts from some of the most renowned astronomers. The name *nebulae* was given to these fuzzy objects that clearly were not stars. Edwin Hubble proved that these objects were indeed other galaxies, much like our own, external to our island of stars, some at colossal distances from us. Hubble (1926) first classified these galaxies according to their apparent morphology into what is now known as the Hubble tuning fork (Figure 1.1). This classification is still present today, and is used to break the population of galaxies down into either *elliptical*, *lenticular*, *spiral* or *irregular* classifications, with the existence of extra features such as bars, dividing objects into subclasses. The different classifications are a reflection of the different properties and forms galactic structures can take, and it has been the role of Astronomers to quantify said properties and

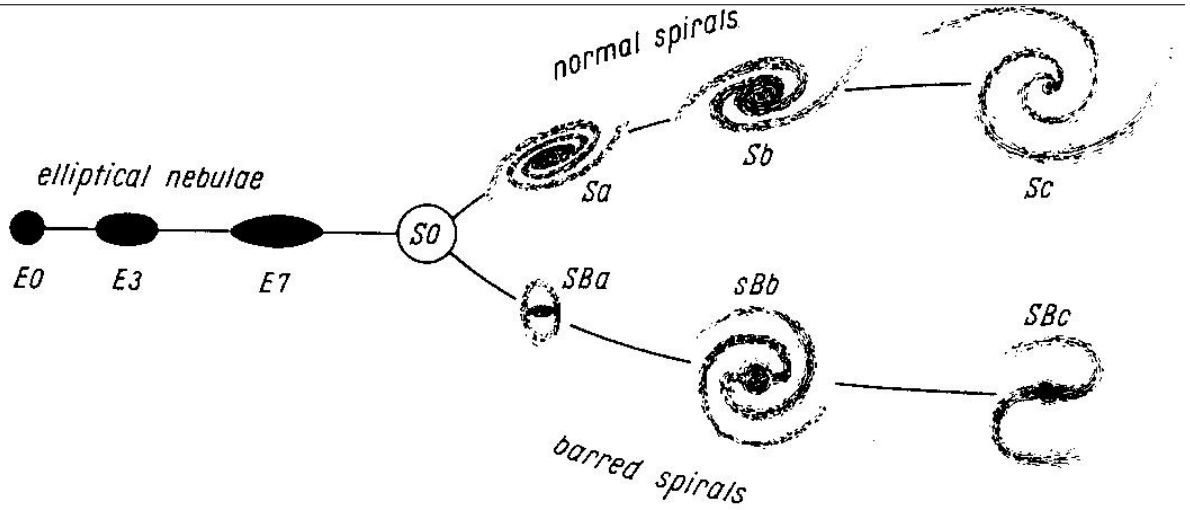


Figure 1.1: The original schematic of Hubble's tuning fork as published in 1936 in his *Realm of the Nebulae*. The elliptical galaxies are presented on the left hand side ranging from E0 - E6, with lenticular galaxies presented as S0. The right hand side shows the spiral galaxies split by the presence of a bar.

understand the physical mechanisms that shape them.

1.2 Galaxy formation: From nothing to something

The following section is an overview of the basic theoretical framework of our current model of how structure forms in the Universe, a logic flow chart is presented (see Figure 1.2) to guide the reader and to highlight the main focus of this thesis.

Galaxy formation models currently favour a cold dark matter cosmology (Blumenthal et al., 1984). The cosmic microwave background is explained successfully with this model (Komatsu et al., 2009), furthermore, there is substantial evidence that the model describes the observed large scale structure of the Universe (Benson, 2010 and references therein).

Structures in the Universe are thought to come about through minute matter density fluctuations, expanded to cosmological scales by inflation (Liddle, 1999). As the dark matter component is entirely governed by gravitational forces, denser regions will collapse, increasing the perturbations. The linear theory of cosmological perturbations provides an accurate description of the early evolution of these perturbations (Shaw & Mota, 2008). Once the perturbation reaches some threshold over-density, it breaks away from the expansion and begins to collapse. This transition signifies non-linear perturbations (Shaw & Mota, 2008).

The consequence of the non-linear, gravitational collapse depends on the denser regions of dark matter and baryonic matter. An approximately stable, near-equilibrium state supported against its

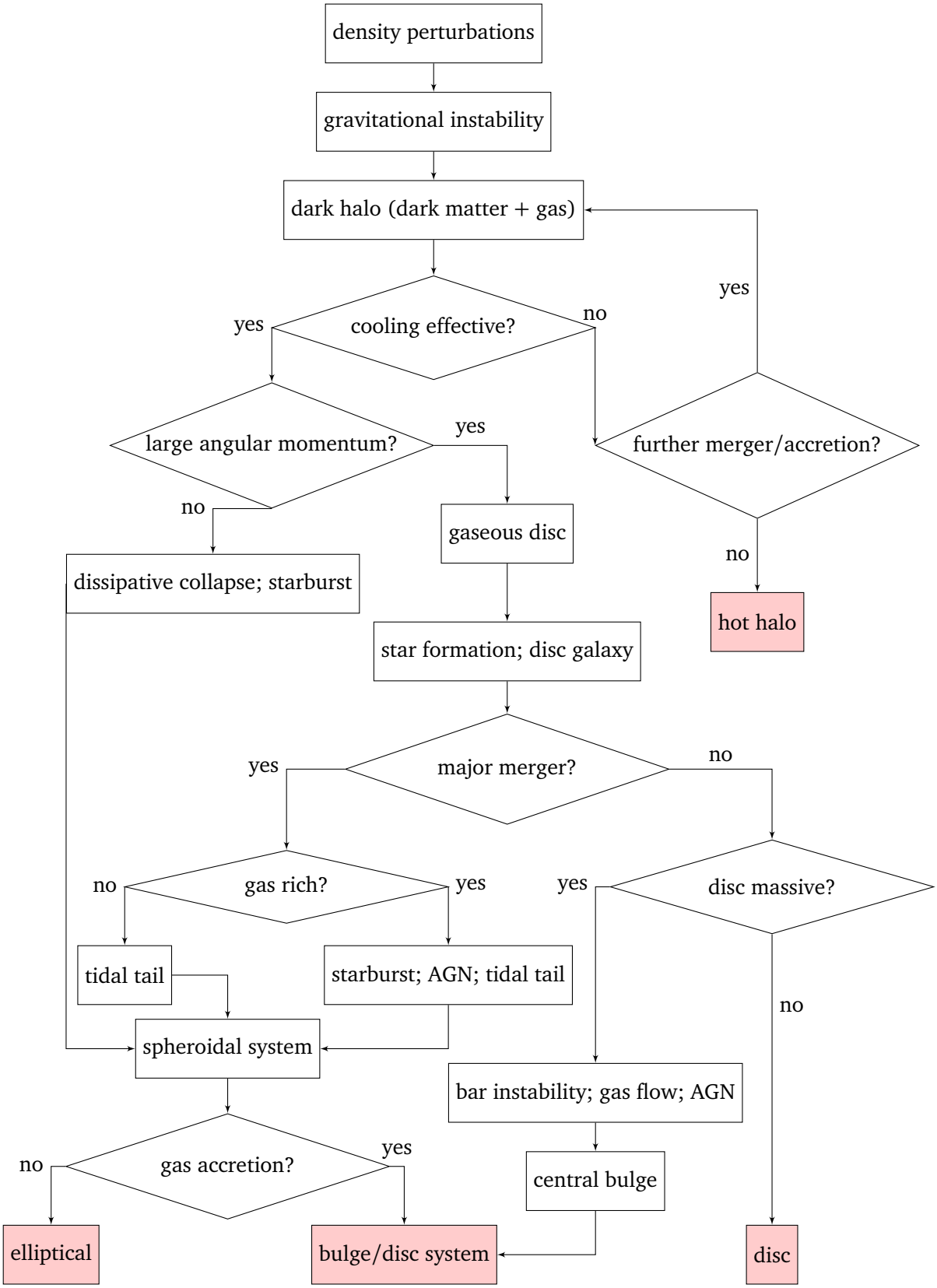


Figure 1.2: A logic flow chart for a Λ Cold Dark Matter model of galaxy formation. This is a very basic idea of galaxy formation and the reader should note that each block and decision marker holds far more detail, and would be near impossible to describe in full. (The flow chart is based on a similar one in Mo et al. 2010.

own self-gravity is the result of this nonlinear evolution of the dark matter density perturbation, known as the dark matter halo. Halo density profiles, shapes, spins and, internal substructure all depend weakly on either mass or cosmology, although the abundance and characteristic density is finely dependent on these (*e.g.,* Eke et al. 1996).

Initially, baryons are thought to be uniformly distributed, tracing the dark matter distribution on scales above the Jeans length (Gnedin & Hui, 1998). Dark matter potential wells are expected to concentrate the force of gravity creating galaxy formation sites. Both the depth of the potential well and the pressure of the baryons will determine how much is accreted into the structure. The cooling of the gas is a vital ingredient to galaxy formation. The overall effect of cooling segregates the baryonic material from the dark matter, and it accumulates as dense, cold gas in the protogalaxy, situated in the centre of the halo. The nature of the cooling process is largely dependent on the atomic species of the gas (different excitation energies) and the initial density. For more insight into cooling with a specific focus into different properties/mechanisms, see the review by Benson (2010) and references therein.

The collapse of baryonic matter is reliant on it being cool, however, this process also leads to the heating of the gas; increased density and temperature leads to the extension of the cooling time. During the collapse the gas may fragment into clumps of small, highly dense cores, finally resulting in the formation of stars. Star formation theory is a broad field in itself and much time and effort has been spent on understanding it (see the review by McKee & Ostriker 2007).

The original concept given to galaxies was the *island* universe. In some cases this can be a reasonable assumption, whereby the galaxy spends most of its time in isolation. Although, in recent decades, evidence of merging galaxies have been observed. The merging together of less massive dark matter halos appears to be the hierarchical nature of structure formation in the Universe, with current high resolution simulations supporting this (Kuhlen et al., 2008). There has also been considerable work in the hierarchical build up of merger trees, which can be thought of as the back-bone to galaxy formation (Jiang & van den Bosch 2014, also see Mo et al. 2010 for more details on the theory behind mergers).

Through the gravitational influence from the dark matter halos, the gas will now have a sufficiently high density to allow a galaxy to form. The shape that the galaxy will have will depend highly on its composition and the properties of its components (*e.g.,* gas and dark matter). To characterise this branch in the galactic evolution tree which involves the internal dynamical effects, the tidal interactions between the dark matter halos and the gas *etc.*, we call this the dynamical evolution. As

the reader will soon find out, this thesis can reside within this field of research. But how do we go from a dense gas cloud within the potential well of a dark matter halo to the galaxies we see today? The following sections will describe in detail the formation of discs, galactic bulges and the various sub-structures within.

1.3 The Stellar Disc

Disc galaxies in general consist of a disc component made up of stars, dust, cold gas (atomic and molecular), a central bulge (discussed more in Section 1.4), a stellar halo, and a dark matter halo. The discs sometimes have very distinctive features such as spiral arms and/or bar components (see Section 1.3.4). A theoretical framework for the galactic disc needs to account for these specific features as well as a variety of observational facts.

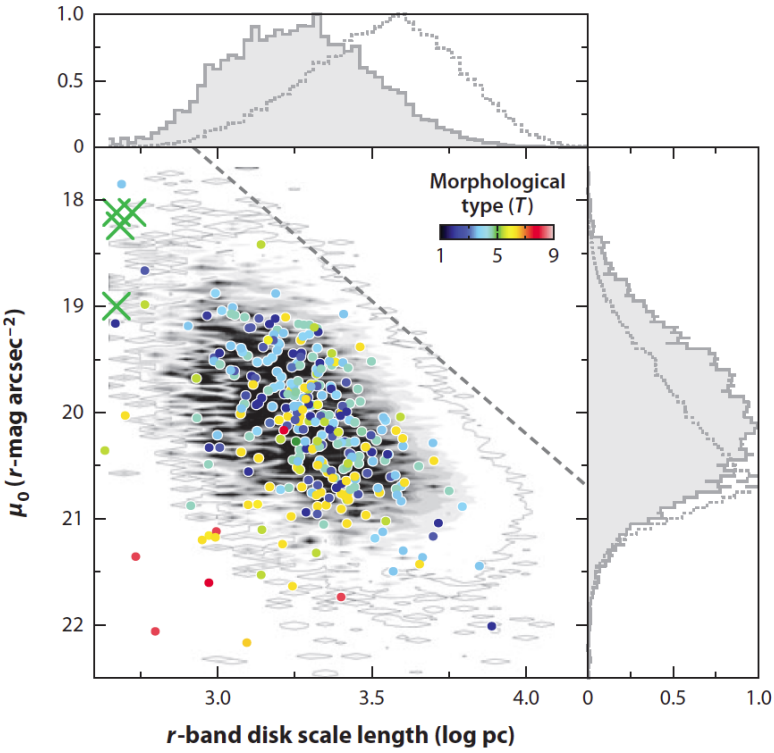


Figure 1.3: The bivariate distribution function of disc galaxy scale lengths and face-on central surface brightnesses from a sample of 30,000 galaxies taken from SDSS, with selection effects corrected. The coloured circles show the 282 classified galaxies with the colour bar indicating the numeric morphological type T with T ranging from Sa ($T = 1$) to Sm ($T = 9$) disc galaxies. The green crosses represent disky ellipticals. A constant disc luminosity with a slope of 2.5 is shown by the grey dashed line. The normalised distribution plots show what is observered (dotted). (From Fathi 2010, adapted in van der Kruit & Freeman 2011)

1.3.1 Observational properties

Observations of disc galaxies show that on average, smaller discs have higher surface brightnesses. One of the most fundamental parameters describing the morphology of disc galaxies is the scale length (h). It describes how the stars within the disc are distributed. Assuming a specific mass-to-light ratio, the scale length can also be used to derive the mass distribution (Fathi et al., 2010). Figure 1.3 taken from Fathi (2010) (adapted in van der Kruit & Freeman 2011) shows the relation between the central surface brightness (μ_0) and the scale length of the disc, suggesting a link between the size of the disc and its brightness.

The radial distribution of surface brightness distribution in the discs of face-on or slightly inclined galaxies can be approximated to be an exponential of the form,

$$I(R) = I_0 \exp(-R/h) \tag{1.1}$$

where $I(R)$ is the intensity as a function of radius, R, in the disc with a central intensity I_0 and scale length, h. Freeman (1970) was the original publication suggesting exponential discs, where B-band observations were used. Freeman (1970) also identified that discs can sometimes differ from the simple exponential profile.

Disc profiles were classified into two groups: type I discs had the simple exponential form, while, type II discs show an outer exponential fall-off, with the inner part sometimes falling below the inward projection of the outer exponential (Freeman, 1970). Type II discs can be identified from the truncation of the stellar population at large radii, roughly 2-4 scale lengths (van der Kruit & Searle, 1981b,a; Pohlen et al., 2002). Using images from face-on galaxies, Pohlen et al. (2002) was able to show that the truncation changes from the shallow exponential of the main disc to the steeper exponential at larger radii. Thus, truncations can be identified as another form of Freeman's type II disc profiles. More recent work, shown through radial surface brightness profiles of early-type galaxies (Erwin et al., 2005) and late-type discs (Pohlen & Trujillo, 2006) presents evidence for a third type of profile. Erwin et al. (2005) classify this type III disc profile as 'anti-truncation': the outer profiles are distinctly shallower in slope than the main disk profile. For a more detailed analysis of galaxy disc properties see the review by van der Kruit & Freeman (2011).

Brighter discs tend to be of a redder colour, although observations show significant scatter. de Jong (1996) presented colour information showing low surface brightness (LSB) galaxies having

bluer colours. The results come about from the mean stellar ages and metallicity and not from the absorption by dust. The work presented in de Jong (1996) show disc galaxies having colour gradients, with the outer regions being bluer than the inner regions.

Another observational quality of bright disc galaxies is that they will be fast rotators (high circular-speed). Discs tend to have flat rotation curves. The circular velocity of massive galaxies tends to rise rapidly at small radii then is almost constant thereafter (Begeman, 1989). Swaters et al. (2000) presented high-resolution Hα rotation curves of LSB galaxies, which were found to have a flat curve out beyond two scale lengths similar to the massive galaxies. The rotation is a direct measure of the gravitational force within the disc, and can be used as a trace of dark matter. The rotation rate of the spirals in the flat part of the circular-speed curve is related to the luminosity of the host galaxies, this is known as the Tully-Fisher relation ($L \propto v_c^{3.5}$) (Tully & Fisher, 1987). The rotation can be measured using a range of techniques, the most common being optical long slit or IFU spectroscopy of HII region emission lines.

1.3.2 High redshift disc observations

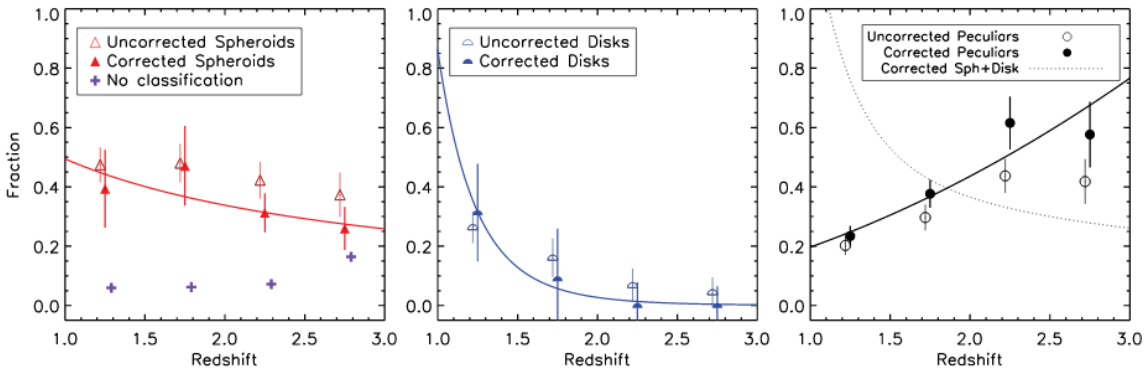


Figure 1.4: The evolution of the fraction of the different galaxy morphologies as a function of redshift from Mortlock et al. (2013). Further analysis has shown that there is a downsizing trend such that the most massive galaxies form into Hubble sequence galaxies earlier than lower mass galaxies Mortlock et al. (2013).

With the advent of the Hubble Deep Fields (HDF) (Williams et al. 1996; Ferguson et al. 2001), the field of high redshift studies has changed drastically since the late 90's. The HDF, and later very deep imaging campaigns, such as the Hubble Deep Field South (Williams et al. 2000), the Great Observatories Origins Deep Survey (GOODS) (Giavalisco et al. 2004), the Hubble Ultra Deep Field (UDF) (Beckwith et al. 2006), the COSMOS field (Scoville et al. 2004), the Extended Growth Strip survey (EGS) (Davis et al. 2007), and the CANDELS survey (Grogin et al. 2011; Koekemoer et al.

2011), have revolutionised the field of galaxy formation and evolution studies, in particular to the study of morphologies.

The critical nature of these deep fields was to obtain photometry at the faintest levels possible in a wide range of bands. Using *photometric redshift* estimation methods (see Dahlem et al. (2013) for a discussion on this method), redshifts for distant galaxies were obtained. These redshifts allowed astronomers to study the evolution over a broad redshift range which was not possible before.

Figure 1.4 shows the evolution of visual morphologies, as defined solely by visual types, from Mortlock et al. (2013) for systems with stellar masses $M_* > 10^{10} M_\odot$. The classification of a spiral, elliptical or peculiar/irregular does not imply that these galaxies have certain properties only that they are visually the same as these classes, as defined in the local Universe. What Figure 1.4 shows is that at $z > 2$ the dominant morphological type is peculiar, while the elliptical and disc galaxies become more populous at lower redshifts (also see Conselice et al. 2005; Kriek et al. 2009; Delgado-Serrano et al. 2010; Mortlock et al. 2013). The Universe is not dominated by the types of morphologies seen in the local Universe until $z \sim 1.9$ (Mortlock et al. 2013). However, galaxies that share similar physical properties with type - such as colours, star forming knots, and tidal features, have not reached the same level as local galaxies at this redshift (Conselice 2014).

How galaxies evolve needs to be addressed with the use of parametric and non-parametric methods. One quantitative measure to observe the evolution of galaxy structures is through the evolution of the Sérsic index, n (Sersic, 1968). The Sérsic model will be discussed in more depth in Chapter 2 but generally n is commonly linked to the shape of a galaxy. Lower n values are associated to more *discy* objects whereas higher n values are connected with more concentrated galaxy structures such as ellipticals. Previous studies of the evolution of derived values of n as a function of redshift show that galaxies have lower n values at higher redshift (*e.g.* Buitrago et al. 2013). Some studies have implied that these galaxies are more disc-like at high redshifts (Bruce et al. 2012), although comparisons to the visual inspection of these systems as well as their physical properties, find they are dissimilar to discs in the local Universe (Conselice et al. 2011; Mortlock et al. 2013). Although these disc-like galaxies have similar light profiles to modern discs they have been observed to be much smaller, have higher stellar masses, and are often undergoing intense star formation making them un-disc-like in regards to the discs observed in the nearby Universe (Shen et al. 2003; Conselice et al. 2011; Buitrago et al. 2013). The question still remains as to how these progenitor galaxies help evolve to form the Hubble sequence we see today.

1.3.3 The origins of the exponential disc

The origin of the exponential stellar disc still remains a challenge in disc formation models. Freeman (1970) specified that the distribution of angular momentum in a self-gravitating exponential disc closely resembles that of a uniformly rotating, uniform sphere. An exponential density distribution with a flat rotation curve also shows similarities (Gunn, 1982; van der Kruit, 1987). To approach this problem, two concepts have been considered; (a) essentially what Freeman (1970) suggested with the surface density distribution of the disc galaxy reflecting the specific angular momentum distribution of the proto-galaxy and (b) the surface density distribution is a result of disc viscosity (Lin & Pringle 1987). Unfortunately both approaches still have their problems. Yoshii & Sommer-Larsen (1989) made an early attempt to describe the formation of the exponential profile. It was shown that an exponential disc could form if the time scales for viscosity and the formation of stars was comparable (also see Slyz et al. 2002 where hydrodynamical simulations of disc galaxies were used to shed new light on the formation and evolution of disc galaxies). It has also been suggested that a disc model with a detailed conservation of angular momentum would give a natural explanation for the exponential profile, potentially forming the truncations as well (Fall & Efstathiou, 1980). However, other non-antisymmetric structures such as bars may redistribute angular momentum, thus it is important to characterise the main properties of these non-axisymmetric structures *e.g.*, spiral arms (Debattista, 2006), in order to shed light on the formation and evolution of discs.

1.3.4 Bars & spiral arms

Stellar bars are quasi-elliptical structures (see Figure 1.5 for an example barred galaxy) which are present in approximately two-thirds of all spiral galaxies in the local Universe (Marinova & Jogee 2007; Menéndez-Delmestre et al. 2007; Aguerri, Méndez-Abreu et al. 2009). Early studies found that the bar fraction evolves significantly (Abraham et al. 1999), while later studies find that bars are already in place by $z \sim 1$ (Elmegreen, Elmegreen & Hirst 2004; Jogee et al. 2004; Marinova & Jogee 2007). It was later shown that the bar fraction increases from $z = 0.84$ to $z = 0.2$, from 20% to $\sim$ 60% of all disc galaxies (Sheth et al., 2008). Sheth et al. (2008) also find that the bar fraction is roughly constant for the most massive and red disc galaxies with lower mass bluer discs showing the most significant evolution.

In many barred spirals, the spiral arms appear to start at the two ends of the bar suggesting that they are tightly related. It is now largely accepted that spiral structures seen in disc galaxies are density perturbations/waves that ripple through the disc structure (Lin & Shu 1964, also described in

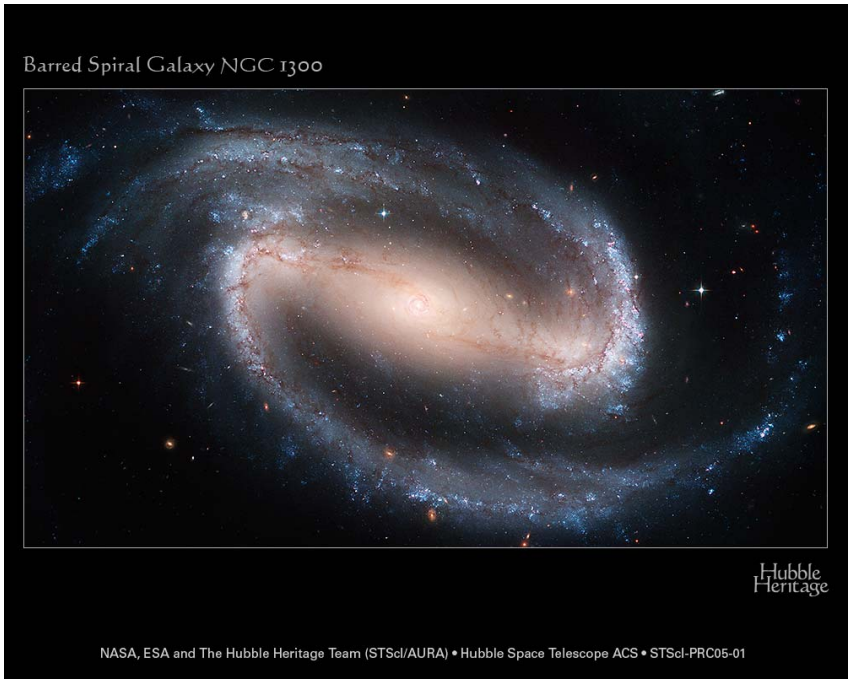


Figure 1.5: The barred spiral galaxy NGC 1300. NGC 1300 has been classified as a SBbc galaxy due to its *grand-design* spiral structure. [NASA and The Hubble Heritage Team (STScI/AURA)]

detail by Bertin & Lin 1996). These density waves can have a great impact on the stellar populations within the galactic components. Numerical simulations show that a significant fraction of stars in galaxy discs undergo large migrations in radius due to resonant scattering from the spiral arms (Roškar et al., 2008a,b). Stars that are initially on circular orbits in the disc, will be scattered into approximately circular orbits with larger radii if they lie near the co-rotation resonance (Roškar et al., 2008b). This can lead to the formation of a steeper exponential decline in the outer disc formed mainly from stars initially located in the central regions of the disc. It is suggested that this process could smooth out the metallicity and age gradients, resulting in a galaxy with older, less metal rich stars at the centre (Roškar et al., 2008b).

Numerical simulations show that the encounter of a disc galaxy with another galaxy can produce a bar-like structure in an otherwise stable disc (Noguchi, 1987). With the disruption of the disc by the bar, a boxy/peanut bulge or a discy bulge may form (Athanassoula, 2008). This is a possible formation scenario for the disc-like bulge (or pseudo-bulge) outlined in Kormendy & Kennicutt (2004). A disc-like bulge is one formed through secular processes in the disc rather than the result of a merger event (see Section 1.4).

1.4 Galactic Bulges

The formation of the bulge component is key when modelling galaxy formation. But what is the galactic bulge? The original meaning comes from the Latin word *bulga* which came to describe the shape of a full bag. The modern Oxford dictionary defines it as a distortion in an otherwise flat surface. Some authors follow this idea and define the bulge as the component that swells out of a disc galaxy when viewed side-on (Figure 1.6 is an image of the Sombrero galaxy, a spiral galaxy seen edge on with a prominent bulge). An early working definition comes from the Carnegie Atlas of Galaxies (Sandage & Bedke, 1994); it states that one of the classification criteria for spiral galaxies is the size of the central amorphous bulge compared to its disc, seen best in nearly edge-on galaxies. Carollo et al. (1999) questions this definition finally concluding that of course bulges are an excess of light (a higher surface brightness) than the inner disc. Another definition follows a more morphological form. Renzini (1999) defines the bulge using the canonical interpretation of the Hubble-Sandage-de Vaucouleurs classification *i.e.,* looking at bulges as ellipticals which have a prominent disc around them, and vice versa for ellipticals that have not maintained a disc during their evolutionary path. Thus defining what a bulge is is no easy task and the nomenclature has changed over the years. With the advent of methods such as photometric decompositions, a more quantitative description was adopted where the excess light that protrudes from an exponential disc quantifying the bulge (Freeman 1970). This photometric definition of a bulge is widely used, usually modelled with a Sérsic profile (Sersic, 1968) separating it from other structures such as discs and bars in multi-component structure modelling (Gadotti 2009; Laurikainen et al. 2010; Méndez-Abreu et al. 2014). Throughout this thesis we use this photometric description of galaxy bulges.

As observations improve, more features are being discovered which make it hard to distinguish the bulge in this way. This leads to astronomers to ask whether bulges of early-type and late-type spirals are really different? Can bulges of different types of galaxies be classified together? (Fathi & Peletier, 2003). Thus, the most operative definition will be to consider the bulge as the extra light in the central region above that of the inner disc. Readers should also recognise limitations in this definition; is it even sensible to consider everything that is in excess of the inner disc to be the bulge? Should we consider the nuclear components of the disc? For the present thesis, it will be sufficient to proceed considering representative properties, leaving details aside until a higher level of understanding is achieved.

Figure 1.6: Sombrero (M104) galaxy. A Sa galaxy hosting a central bulge [NASA and The Hubble Heritage Team (STScI/AURA)]

1.4.1 Observational properties

The advance of optical (CCD) and near-infra-red instruments in the late-80s and 90s meant that full 2D data modelling was achievable which subsequently lead to a quick increase in the quality of bulge photometry. Reliable photometric models for the bulge were then subtracted to understand individual components of galaxies. The general features of bulges can now be described in a variety of ways.

• Photometry: A morphological point of view

de Vaucouleurs (1948) produced the quantitative description of elliptical galaxies which was expanded to bulges of disc galaxies, including the surface brightness profiles following the radius as $r^{1/4}$. Sersic (1968) improved this relation by replacing the 1/4 in the $r^{1/4}$ law with $1/n$, where n is fitted to the data. Kent (1985) measured the shapes of bulges as well as their radial profiles and concluded that the ellipticity shows little correlation with the effective surface brightness of the bulge (see Figure 1.7). Kent (1985) stated that if all the bulges followed the same law, *i.e.,* they were all oblate spheroids viewed from different directions, more edge on galaxies would have a larger ellipticity thus they would have a higher surface brightness. The absence of this correlation suggested that bulges must have more complex shapes than simple oblate spheroids. This has been later confirmed due to the triaxiality of bulges (Méndez-Abreu et al., 2010). Andredakis et al. (1995), using the two-dimensional generalization of Kent's model-independent photometric

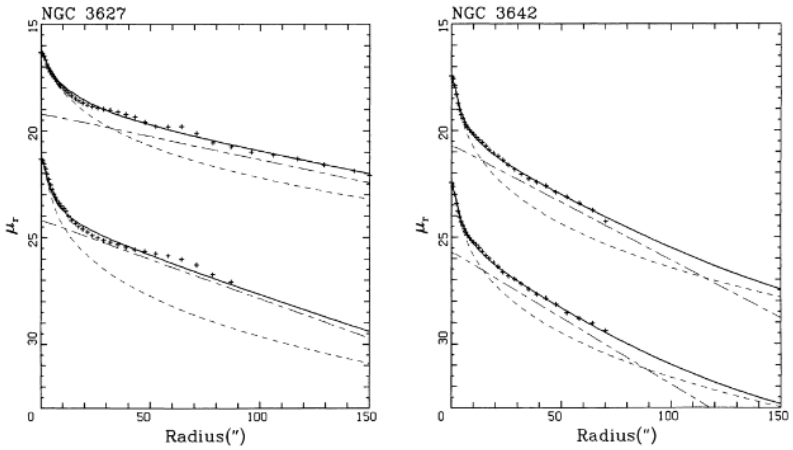


Figure 1.7: two-component model fits to the observed profiles of two S0 galaxies from Kent (1985) were some of the first photometric fits to bulges were done. The solid line is the sum of the bulge and the disc with bulge and the disc components being represented by the short-dashed and the long-dashed lines respectively.

decomposition method for extracting K-band light profiles, fitting with the more general Sèrsic profile, found early-type galaxies were better fit with a de Vaucouleurs profile while bulges in late type galaxies had more of an exponential profile.

The $r^{1/4}$ law was still used in bulge-disc decompositions as it was favoured for high redshift (Simard et al., 2002) and early-type bulges. It was during this time that the use of the Sèrsic profile began to seem more and more relevant in bulge-disc decompositions. Several studies including early-type bulges (Khosroshahi et al., 2000; Graham, 2001) found that the bulge luminosity scaled with the Sèrsic shape indices (from $n < 1$ to $n > 4$). The central regions of the profile dominate at higher values of the Sèrsic index. Bringing together knowledge of the inner regions of the disc (Phillips et al., 1996), compact sources (Rest et al., 2001; Pizzella et al., 2002), and star formation (Carollo et al., 2002), present in the galaxy centres, we begin to understand the difficulty in disentangling the different light profiles. The sources present in the central regions contribute to the light resulting in a higher n value. For example, Balcells et al. (2003) used the *Hubble Space Telescope* (HST) near-infrared imaging instruments, to measure the surface brightness profiles of spiral (S0-Sbc) galaxies at high resolution. The result found that nuclear point sources were blending with the light of the bulge due to the point-source properties, namely the point spread function (PSF) (Balcells et al., 2003).

• Kinematics

The study of motions of stars in galaxies reveals a wealth of dynamical information about these systems. Due to the great distances to external galaxies, we cannot obtain all the data required to

determine the distribution of the stellar component within. Also obscuration due to dust further complicates the analytical process as well as the mixing of stars making it a complex process in separating different stellar components. Only line-of-sight velocities and angular coordinates are observable. Until the seminal work of Kormendy & Illingworth (1982), astronomers held the traditional picture of bulges being elliptical galaxies with a disc. This idea was motivated by the very similar properties observed for bulges and ellipticals, i.e., morphologies, stellar content, and the relation between the luminosity and the central velocity dispersion (also known as the Faber-Jackson relation, Faber & Jackson 1976). However, the bulges for which kinematics were available were all situated in early-type spirals with large central regions. Kormendy & Illingworth (1982) studied the relation between the ratio of the rotation velocity, V, and random motions, *i.e,* velocity dispersion, σ, with the ellipticity, ϵ, of the galaxy. Evidence was provided that bulges follow the oblate-spheroid picture. Kormendy (1982) then added to the previous list of intermediate wound spiral (S0-Sb) bulges with a set of lens-shaped barred spirals (SB0) bulges. All the SB bulges were found to be rotating at least as rapidly as the oblate-spheroids, with some rotating faster than previous estimates for an isotropic rotator. This suggested that these systems contained more disc-like kinematics than spheroidal.

The study of stellar kinematics took a giant leap with the discovery of thin, rotating discs dominating the light in the inner parts of some galaxies, accompanied by a local minimum in the velocity dispersion (Emsellem, 1998). Long slit spectroscopic observations of spiral galaxies show stellar velocity dispersions (in the radial direction) were decreasing exponentially with radius (Bottema, 1989, 1993). With the integral-field spectrograph SAURON operating on the William Herschel Telescope, Falcón-Barroso et al. (2006) was able to characterise the central local minima in the velocity dispersions in some 13 Sa and Sab galaxies. The star formation rates (SFRs) in the sample were similar to what was seen in normal disc galaxies, with the low velocity dispersions being linked to regions of star formation. Evolution of the galaxy over a period of time similar to the dynamical time of the galaxy could result in this velocity dispersion drop due to gas falling into the centres; this process is discussed further in Section 1.4.3 on the formation theories of bulges.

• Stellar populations

Early population synthesis studies were performed with libraries of stellar spectra or libraries containing stars and globular clusters. It wasn't until Bica (1988) used a base of star cluster integrated spectra to undertake a two parameter analysis composed of age and metallicity. From the age and metallicity an idea of the bulge stellar population can be formulated.

Following the pioneering effort of Bica (1988), it soon became apparent that a metallicity-luminosity (Z/L) relation was present (Jablonka et al., 1996; Idiart et al., 1996). Spectroscopic studies in the late nineties provided evidence for similarities in the metallicities of both spiral and elliptical galaxies (Jablonka et al., 1996; Idiart et al., 1996). Similarities were held in the mean metallicity, with ellipticals observed to have an Iron to Hydrogen ratio being roughly half that of bulges. On the other hand, the α elements to Iron ratio, $[\alpha/Fe]$, was approximately equal to one another. The stellar populations were suggested to scale with the central velocity dispersion and the bulge luminosity rather than the classification of the galaxy (Jablonka et al., 1996). Comparisons were later made between M31 and the Galactic bulge (Jablonka et al., 1998). 19 globular clusters present in the central regions were found to be mostly metal-poor which was proclaimed to be similar to the Galactic bulge. However, two of the 19 clusters observed were highly metal-rich and compared well with the nuclei of giant ellipticals and the semi-stellar nucleus of M31 itself, these metal-rich clusters in the core were deemed rare. Hammer et al. (2001) speculated that luminous compact galaxies (LCGs) were instead the progenitors of bulges of massive spirals. Spectroscopic observations of a sample of LCGs were compared to present day bulges of massive spirals. Similarities were found in their stellar masses, metal abundances, along with evidence to suggest low surface brightness components around the high surface brightness cores. While similarities maybe present, the statement that LCGs are the progenitors to modern day bulges is a speculation (more on the evolution scenarios will be discussed in Section 1.4.3).

Photometric studies of bulges situated in early-type galaxies show that they differ from bulges in late-type spirals (Falcón-Barroso et al., 2002). Using HST, Falcón-Barroso et al. (2002) has showed that bulges of later morphological types have slightly younger ages. The analysis of stellar populations through spectroscopic observations of edge-on spiral galaxies spanning the Hubble sequence show the vast majority of bulges are of an older age, lower metallicity and higher $[\alpha/Fe]$ in the outer region than the central parts (Gorgas et al., 2007; Jablonka et al., 2007). Jablonka et al. (2007) suggest that the effective radii and the Hubble type of the parent galaxies play a minor role in causing spatial variations in the stellar populations.

1.4.2 High redshift bulges

Studying the high-redshift progenitors to local galaxies can provide us with invaluable insights into the key mechanisms that drive the evolution of these galaxies, constraining parameters for the use in formation models. Data from local galaxies cannot distinguish between the various evolutionary scenarios for bulge, including internal processes triggered by the presence of bars in the host galaxy

(Kormendy & Kennicutt, 2004; Combes, 2007).

In order to answer the questions surrounding bulge formation, observations of galaxies at intermediate redshifts will help in determining parameter estimates. Advances in deep multicolour imaging from HST provides a glimpse into the photometric properties. Ellis et al. (2001) observed that the colours of the bulges of spiral galaxies are redder than ellipticals at the same redshift. Analysing the internal optical colours of early-type and bulges in spirals with $I_{AB} < 24$ magnitudes in the northern and southern Hubble Deep Fields (HDF), Ellis et al. (2001) determined a diversity in bulge colours over the redshift range $0.3 < z < 1$. At higher redshifts, bulges were found to be as red as the reddest ellipticals at similar redshifts in the HDF. They concluded that periodic bursts of rejuvenation may occur within the observed bulges as the stellar mass grew by 15-30% since $z \simeq 1$. These results were challenged by Koo et al. (2005), who used HST to separate bulge-like and disc-like components of 71 luminous galaxies ($I_{AB} < 24$ magnitudes) in the rich cluster MS 1054-03. Their key findings were that the luminous bulge components were much redder than had previously been measured by $U - B \sim 0.45$. Koo et al. (2005) also found that 85% of their field sample had bulge colours at $z \simeq 0.8$, as red as high redshift E/S0 galaxies.

With the discrepancies in photometric studies, MacArthur et al. (2008) argue that photometry and colors are not sufficient to resolve the issues surrounding the understanding of bulge evolution. They used spectroscopic observations of galaxies in the GOODS field to study the diversity and evolutionary history of distant galactic bulges. The sample of 137 spiral galaxy bulges within the redshift range $0.1 < z < 1.2$, show a broader range of optical colours than that expected for passively evolving populations. MacArthur et al. (2008) also estimated bulge mass from stellar velocity dispersions using spectroscopic observations from the Keck II telescope (see Figure 1.8). More massive bulges were found to be as old and as red as massive ellipticals, however smaller bulges were found to have diverse star formation histories (SFHs), with significant star formation below $z \sim 1$.

Recent studies have added to the understanding of the evolution of galactic bulges. Bruce et al. (2012) isolated a sample of 200 very bright (< 27 mags in the HST WFC3 H_{160}-band) very massive ($M_* > 10^{11} M_\odot$) galaxies in the CANDELS-UDS field with photometric redshifts in the range $1 < z < 3$. A detailed analysis of their rest-frame optical morphologies, and how these vary with redshift, mass and SFH was undertaken. The bulges within their sample show evidence for a growing bimodality in the size-mass relation with increasing redshift. The fraction of bulges consistent with the local mass-size relation increases from $15 \pm 9\%$ at $2 < z < 3$ to $20 \pm 5\%$ at $1 < z < 2$.

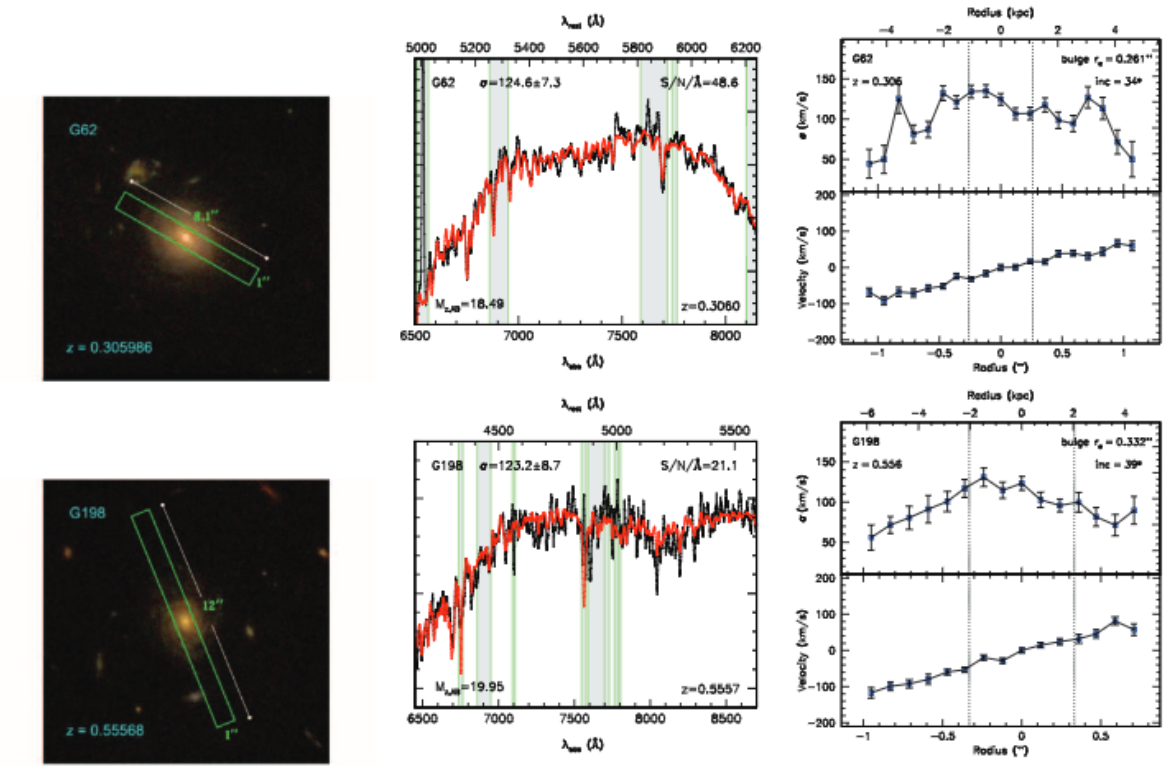


Figure 1.8: Spectroscopic measurements for two bulge-disc galaxies from MacArthur et al. (2008). *Left panels*: Color ACS images and slit orientations. *Centre panels*: Observed central galaxy spectrum (black lines) and the best fitting stellar template (red lines) convolved to the measured velocity dispersion. *Right panels*: Kinematic profiles. *Top*: Velocity dispersion as a function of the light-weighted radius. *Bottom*: Radial velocity profiles (i.e. rotation curves), shifted to zero velocity at the centre. The vertical dashed lines indicate the effective radii for both the bulge and the disc.

The trends suggest that the lower envelope of galaxy size is a function of mass. The morphologies in the sample dramatically change with redshift, with hints that $z \simeq 2$ is the key epoch of transition. Low redshift galaxies are found to be more bulge dominated, while between $1 < z < 2$ there are predominately increasing bulge-disc systems. A $z > 2$, the galaxies are found to be more disc dominated. Bruce et al. (2014) extend this analysis by conducting individual component spectral energy distribution (SED) modelling to estimate the bulge and disc stellar-mass and star-formation rates (SFRs). They support the previous results that at higher redshifts, the galaxies are more disc dominated with the galaxies becoming more bulge dominated with decreasing redshift. However, it is concluded that the fractions of passive galaxies being disc-dominated and star-forming galaxies being bulge-dominated is significantly less than what was previously suggested by past studies at similar masses and redshifts. Although most of the massive galaxies may undergo major mergers which both quench their star formation and change the morphologies, however, they are not the only physical process quenching star formation, leaving a massive disc intact; active galactic nuclei, halo

quenching and violent disc instabilities may also play a role (See Section 1.4.3 for further details).

1.4.3 Bulge classifications and formation theories

In the previous sections we have discussed the central regions of galaxies as if they all share a common evolutionary path. This is however untrue, as observations of bulges have led to the identification of different schematical classifications: classical type, boxy/peanut and disc-like (or pseudo) bulges (Athanassoula, 2005). The emphasis on classifying bulges is so that we can begin to understand the different formation histories and relate these with the observations of bulges today. However, the picture is more complex than simple paths from one type to another. The Universe has been constantly evolving, with dynamical characteristics in galaxies changing as we look back to earlier cosmic epochs (see Figure 1.9). To complicate things further, it has been suggested that different types of bulges can reside within one another in the same galaxy (Athanassoula, 2005). How might this arise? The mechanisms that are thought to transform these galaxies can be considered either internal (secular evolution, see Kormendy & Kennicutt 2004) or external (galaxy mergers). Secular evolution is concerned with how the galaxy may change over a long period of time, this involves the formation of a bar and its interaction with the disc component. On the other hand, the hierarchical build up of smaller sub-structures is suggested to be the key process that forms classical type bulges we see in the local Universe. This section attempts to bridge the gap between the observations of the different types and the suggested theoretical ideas that describe the formation of galaxy bulges.

• **Classical bulge**

The merging of objects along with the accretion of smaller substructures through the lifetime of the galaxy is thought to cause the formation of classical bulges (Kauffmann, 1996; Baugh et al., 1996; Cole et al., 2000). Observational similarities between elliptical galaxies and classical bulges have led some authors to think of bulges as being ellipticals with surrounding discs (See Figure 1.6). There have been suggestions that the classical bulge may have formed around the same time as the galaxy disc, possibly shortly before (Sommer-Larsen et al., 2003). Hydrodynamical simulations have shown that SF within the galaxy could tip the galaxy into being more disc dominated or bulge dominated. This is due to the energy released in the process of forming stars transferring the angular momentum that would normally create clumps of gas and stars into discs (Debattista, 2006).

Classical bulges do share some fundamental properties with elliptical galaxies, for example they both have surface brightness profiles that have high Sèrsic index (Wyse et al., 1997). Also their kinematic properties are well represented by dynamical models of rotating flat spheroids with insufficient

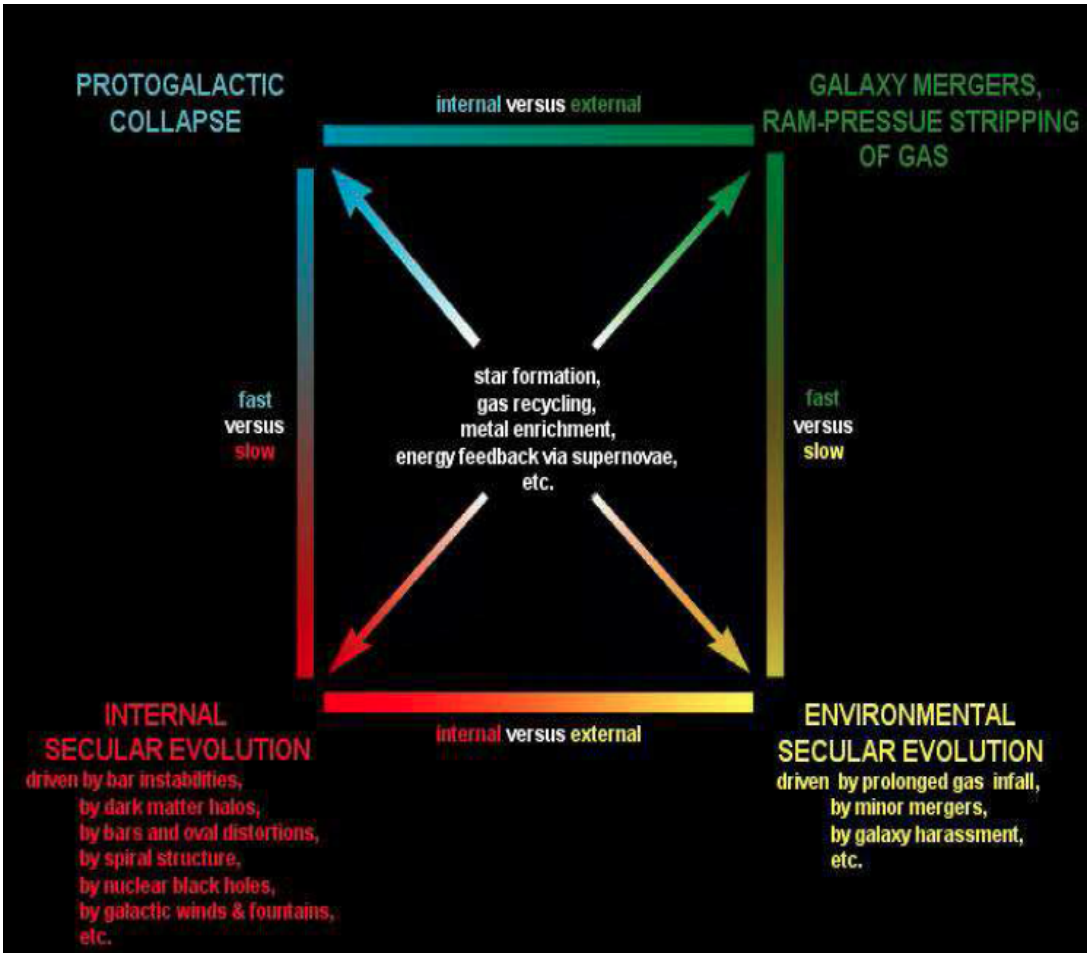


Figure 1.9: Diagram depicting the different evolutionary paths a galaxy may take, affecting the formation of a bulge. (Taken from Kormendy & Kennicutt 2004)

anisotropy (Davis et al. 1983). Classical central regions mainly follow a $r^{1/4}$ luminosity profile with the stellar populations often being older and less metal rich (Jablonka et al. 1996). Studies have shown that the bulk of the stellar populations of these structures formed between $3 < z < 5$ (Wyse et al., 1997; Mehlert et al., 2003; Thomas et al., 2005). The kinematic properties of these systems tend to be dominated by random motions similar to elliptical galaxies, thus hinting at a common evolutionary path (Bender et al. 1992).

- **Boxy/peanut bulge**

Boxy/peanut (B/P) bulges owe their name to the observations of side-on disc galaxies (Figure 1.10). The modern interpretation of these structures refers to secular evolution, involving the creation and interplay of a galaxy with its bar (Martinez-Valpuesta et al., 2006). As the bar forms, the stellar component moves in the perpendicular direction with reference to the disc plane, due to the coherent bending of the bar (see Athanassoula 2016 and references therein). The B/P structure is thus present

Figure 1.10: Edge-on view of a Sb galaxy with a boxy bulge.

in observations of edge-on galaxies with bars. Therefore, these B/P systems have the same kinematic and photometric properties of the bars. The bar morphology also shares specific characteristics with the B/P bulge while the stellar population shows similarities to the inner disc (Combes et al. 1990). Observation have also linked B/P bulges to the bar with the strength of the bar being directly proportional to the protruding of the B/P bulge (Lutticke et al. 2000).

The B/P bulge will be masked onto the central region with a low inclined galaxy making comparisons with side-on structures difficult. Debattista et al. (2005) described a novel kinematic technique to distinguish B/P bulges in face-on galaxies using a double minimum in the forth order moment of the line of sight velocity distribution along the major-axes of the bar. Minima occur at the location where the stellar velocity density distribution plateaus. This method was observationally verified by Méndez-Abreu et al. (2008b) leading to the identification of a peanut bulge in the face on galaxy NGC 98. Identification of B/P bulges located in bar structure gives new insight into the relation between these components.

• Disc-like bulge

Early epochs of galaxy formation are dominated by violent collisions between neighbouring galaxies becoming the foundations of modern day classical bulges and elliptical galaxies. Since the rate of mergers has declined since $z \sim 1.5 - 2.5$ (Conselice et al. 2008), secular processes have been the

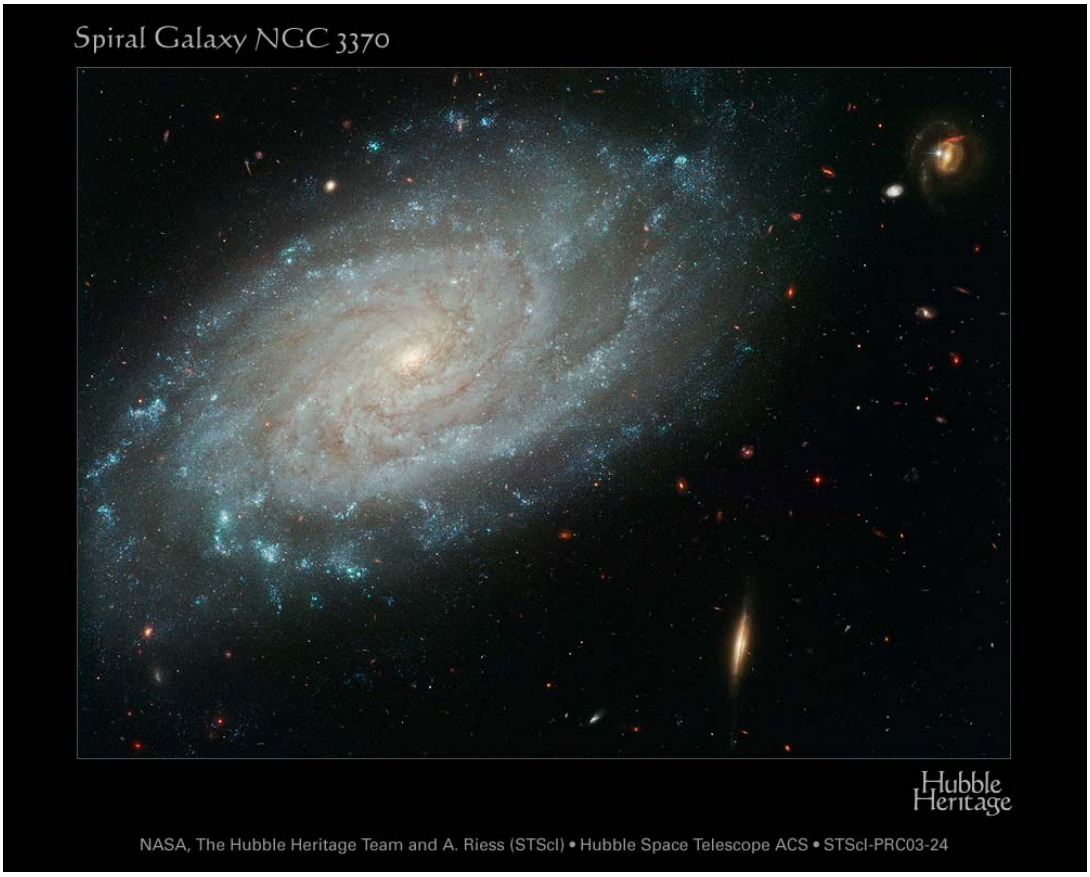


Figure 1.11: NGC 3370 galaxy. A Sc galaxy hosting a disc-like bulge. [NASA and The Hubble Heritage Tam (STScI/AURA)]

dominant mechanisms that shape galaxy structures (Kormendy & Kennicutt, 2004). We discuss two possible pathways proposed, for forming bulges through secular process:

Hypothesis one: Gas is funneled by the gravitational potential of the bar, which then accumulates in the galaxy centre forming the inner disc. Once the inner component reaches a critical mass it starts forming stars and thus the discy-bulge (see Figure 1.11 for an example of a galaxy with a disc-like bulge) (Athanassoula, 1992; Heller & Shlosman, 1994).

Hypothesis two: If the accumulation of material in the galaxy centre becomes too massive the bar could be destroyed due to the instabilities it produces. Thus the disc-like bulge will be a mixture of bar and inner stellar components with a flattened structure, following an exponential luminosity profile, rotate as fast as discs, and host a young stellar population (Hasan et al., 1993; Shen & Sellwood, 2004)

• **Mixed types**

A combination of classic- and disc- bulge structures have been found to occur in the same galaxy

(Athanassoula, 2005). Furthermore, B/P can also be combined with the different types. The exact number of these types of systems still remains uncertain, with various combinations successfully describing observations. The bulge component is not the only structure that can be present in the central regions of disc galaxies: nuclear stellar discs (characterised by small scale lengths and higher central surface brightnesses compared to smaller discs (Morelli et al., 2010)), nuclear clusters (Seth et al., 2006), and nuclear bars (subkiloparsec scale bars (Erwin, 2004)). With the increasing spatial resolution of our telescopes, we are beginning to understand that the central regions are far more complex, containing several structural and kinematic components formed through a complex mixture of scenarios.

1.5 Conclusion & outlook

The full story of how galaxies form and evolve remains yet to be answered in Astronomy. Distinguishing which evolutionary scenario that hints the galaxies we observe in the local universe to those at early epochs of creation requires a detailed study of their principal structural components (bulges, bars, and discs) using a multidisciplinary strategy (structure, kinematics, and stellar populations), with comparison to state of the art simulations.

With ever improving observations as well as a better understanding of the theory, our understanding of galaxy formation is continuously updated. Recent publication have supported the existence of blue elliptical (Huertas-Company et al., 2010), and red spirals (Masters et al., 2010) which contradict the standard classification of galaxies into passive early-type and star-forming late-type. Although detailed studies of galaxies have been conducted since the 1920s, the recent appearance of publications challenging past ideas subsequently supports the need for fresh new studies.

There are many different techniques that attempt to peer into the underlying physical process responsible for galaxy morphologies. Two observational approaches that explore the formation mechanisms and evolution of local bulges are: (i) the study of the imprinted fossil records of bulge observations, or (ii) using cosmology as a time-machine to look back at the early galaxies to reveal their evolution and even perhaps the formation of bulges in-situ. With the advent of large and deep surveys (GOODS, COSMOS, ..), detailed studies of galactic bulges at large cosmic distances can now be attempted. Some of the remaining questions to be answered are:

- How many of the bulge properties reflect external hierarchical build up or secular evolution?
- How, or even if, transformations in morphology are linked to star-formation history?

- What is the prominent bulge type at different redshifts?
- What are the descendents of high-redshift bulges?
- What is the connection, in a cosmological framework, between bars and bulges?
- What is the relative contribution of each bulge type to the total bulge within any individual galaxy?
- How do the properties of different bulge types relate to those of their host galaxies?

2

A Bayesian approach to 2D Photometric Decompositions: Methodology

2.1 Introduction

The long history of studying galaxy structures pre-dates even when we knew galaxies were extra-galactic. The extended and resolved morphology of galaxies separated them from stars in the night sky. Today measuring the structure of galaxies is one of the most common ways we describe and study these objects. At first this involved visual inspections of photographic images and now has expanded to involve a whole host of quantitative methods that measure galaxy structures from the local to very distant Universe.

Here I describe the various ways in which galaxy structure is measured and quantified for comparisons across cosmic time. For high redshift structural studies it is important that they take general features that can be measured and compared to galaxies in the local Universe. For nearby galaxies there is a great wealth of physical properties that can be observed that cannot be examined at high redshift. Thus the majority of previous works have focused on three broad methods to mea-

sure structure: visual/machine learning classifications, non-parametric and parametric methods. We now describe each method individually with the most focus placed on parametric procedures as this will be the main method used in this thesis.

2.1.1 Visual morphology and machine learning methods

The human brain is extremely powerful at pattern recognition so it was only natural that Hubble (Hubble, 1926) started to spot similarities between the objects and began to classify them. This major system of classification was largely developed through Hubble (1926), de Vaucouleurs (1956) and Sandage (1961, 1975) as outlined briefly in the introduction (see the review by Buta (2013) for a detailed overview into visual classification). When studying the morphologies of distant galaxies the visual classifications can only be placed into a few limited and well defined classes: spirals, ellipticals, and irregulars/peculiars. As outlined in the introduction, spirals can be further classified according to the presence of a bar or not.

Today visual classifications are still widely used but have been revitalised by the World Wide Web in connecting galaxy classifiers across the globe to collect vast amounts of galaxy structural information (Galaxy Zoo; see Lintott et al., 2008, 2011; Willett et al., 2013; Simpson et al., 2014; Simmons et al., 2016). Visual morphological classifications have been performed on hundreds of thousands of galaxies using available large surveys such as the Sloan Digital Sky Survey (SDSS) and the Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey (CANDELS) (Lisker, 2008). Another issue is that high redshift galaxies that look elliptical or discy do not have the same characteristics as systems with the same morphologies seen in the nearby Universe (Mortlock et al., 2013; Buitrago et al., 2013).

With the advent of powerful computing resources such as Graphic Processor Units (GPUs) it has became possible to mimic the human brain (e.g., deep learning) in how it captures the full, complex distribution of light, further advancing this method of classification (Huertas-Company et al., 2008; Dieleman et al., 2015). Huertas-Company et al. (2015) trained an algorithm with visual morphologies available in GOODS-S and then applied it to the remaining fields in CANDLES. Galaxies were assigned a frequency value that resembles what a hypothetical classifier in a visual classification scheme (like Galaxy Zoo) would have given to the galaxy. This essentially gave each galaxy a probability of being either an elliptical, disc, irregular or point source. The same limitations to visual classifications could be said to also play a role in high redshift studies using machine learning techniques, however, due to the nature of this method large arrays of parameters (from

parametric and non-parametric methods alike) can be combined in order to better classify objects (Huertas-Company et al., 2008, 2015).

2.1.2 Non-parametric measurements of structure

The quantification of different structures has historically advanced our understanding of galaxy structures. Describing the structure of a galaxy in a quantitative manner has the advantage over qualitative morphology estimates due to its reproducibility. Quantitative classifiers can either be parametric or non-parametric.

Non-parametric methods of measuring the structures of galaxies were first performed by Morgan (1962) with attempts to quantify the light concentration in galaxies. However, it was not until the mid-1990s that extensive non-parametric measures were done (Abraham et al. 1994, 1996; Shade et al. 1995, Rix & Zaritsky 1995; Conselice 1997). These early papers show that quantitative galaxy structure correlates with other parameters such as colour and peculiar features indicating galaxy interactions i.e. mergers (*e.g.* Rix & Zaritsky 1995; Conselice 1997).

The present methods for measuring structure through non-parametric classifiers is the CAS system (*e.g.* Conselice 2003) and through similar quantitative measures (Takamiya 1999; Papovich et al. 2003, 2005; Abraham et al. 1994, 2003; Lotz et al. 2004; Law et al. 2007; Freeman et al. 2013). These measurements are designed to capture the underling make-up of a galaxy without any prior assumptions. There are typically four non-parametric indices that are commonly used in literature: the Petrosian radius, the asymmetry index, the light concentration and the *clumpiness* index.

The Petrosian radius (Petrosian 1976) is described as the location where the ratio of surface brightness at a radius, $I(R)$, divided by the surface brightness within the radius $< I(< R) >$, reaches some threshold value, $\eta(R)$ (typically when $\eta(R) = 0.2$) (Bershady et al. 2000; Conselice 2003; Lotz et al. 2004). The asymmetry index (A) measures how asymmetric a galaxy is after rotating along the line of sight centre axis of the galaxy by 180 degrees. In general terms it is a measure of what fraction of the light in a galaxy is in non-symmetric components. It is often used to identify merging galaxies (Conselice 2003). The concentration index (C) quantifies how much light is in the centre of a galaxy compared to the outer parts. This index has been shown to correlate with the Sérsic index (see Section 2.2.1 for a description on the Sérsic light model) (Graham et al. 2005). One limitation with C is that it is very dependent on where the measurement is taken as different regions and different radii can produce very different values (Graham et al. 2001, 2005). The final most common non-parametric measurement is the clumpiness or smoothness (S) index which describes

the fraction of light in a galaxy which is contained in clumpy distributions (Isserstedt & Schindler, 1986; Takamiya 1999; Conselice 2003). A common way to measure S is to subtract a blurred image of a galaxy from the original galaxy image, leaving a residual map containing irregularities. A drawback of this measurement is that the centres of the galaxy are usually removed as they often contain unresolved high-spatial frequency light.

2.1.3 Parametric measurements of structure

Initially galaxy structure was quantified through the use of integrated light profiles. This allowed more quantitative classification schemes based on a parametric description of a galaxy's structure. These quantitative methods have two major advantages: 1) they are reproducible, and 2) it is easy to understand biases with the use of simulations. These parametric descriptions of structure can be applied to large samples of galaxies to better recover reliable information regarding their evolution over time. A historical account of the methods most commonly used in literature will now be discussed.

One-dimensional photometric decompositions

The earliest efforts of galaxy structural characterisation concentrated on simple one-dimensional (1D) intensity profile fitting. These profiles are measured by taking the average intensity of a galaxy at a given radius, and then determining how this intensity changes as a function of radius. This lead to the early discovery of the de Vaucouleurs profile (de Vaucouleurs, 1948) for describing elliptical galaxies and then the realisation that disc galaxies follow a tight exponential profile (de Vaucouleurs, 1956). The first attempts in photometric decompositions of images of galaxies came from Freeman (1970) where parametric methods were used to describe the luminosity of exponential discs. The method consisted of modelling the galaxy surface brightness dominated by the disc with a linear fit in magnitude units.

Increasingly complicated models came in the works by Kormendy (1977) and Kent (1985). Kormendy (1977) was the first to address the need to fit different sub-structures within a galaxy with different models. First the disc would be subtracted from the total surface-brightness of the galaxy assuming an exponential profile. Then, the bulge component was fit with a de Vaucouleurs profile ($R^{1/4}$), with this then being subtracted from the total surface brightness in an iterative process. One major draw back of this method is the bias made when subtracting the disc first, as the bulge light is likely to partially mix into the disc. Kormendy (1977) then introduced a modification to this method in that both components are fit at the same time using a non-linear least-squares pro-

cess. This method is still used today when dealing with photometric decompositions of multiple components.

Boroson (1981) introduced a method where photometric decompositions are based on ellipticity averaging surface-brightness profiles and modelling different galaxy components with 1D profiles. It is based on the extraction of the surface-brightness profile along axes with different position angles, all of them decomposed as if they are independent. However, these 1D profile fitting methods have one major limitation i.e. they do not take into account the intrinsic shapes (Prieto et al. 2001) or position angle (Trujilo et al. 2001) of the bulge or disc, producing systematic errors in the results (Byun & Freeman 1995).

Two-dimensional photometric decompositions

The method improved from 1D to fit all of the pixels within an image in two-dimensions (2D) to include the intrinsic shapes and position angles. Photometric decompositions have since been done this way and there have been a number of algorithms developed since; GIMP2D (Simard, 1998), GALFIT (Peng et al., 2002), BUDDA (de Souza et al., 2004), GASPHOT (Pignatelli et al., 2006), GASP2D (Méndez-Abreu et al., 2008a), IMFIT (Erwin 2014), GALPHAT (Yoon el at. 2011) and PROFIT (Robotham et al. 2017). Each algorithm was designed to solve different problems of galaxy decompositions with each using differing functions to parametrise the galaxy components and using various minimisation or probabilistic techniques to fit the models to the data.

Scorza & Bender (1990) developed a process that resembled the method first used in Kormendy (1977) where a model image was created for the disc component using an exponential profile. This image was then subtracted from the real galaxy image forming a residual map containing the bulge light. The best fitting disc is then found by comparing the bulge isotopes with perfect ellipses. Another 2D image fitting method is the Multi-Gaussian Expansions (MGE) (Bendinelli 1991; Monnet et al. 1992; Emsellem et al. 1999; Cappellari 2002). The method involves expanding the galaxy surface-brightness distribution in a series of Gaussians given by

$$\Sigma(R,\theta) = \sum_{j=1}^{N} \frac{L_j}{2\pi\sigma_j^2 q_j} \exp\left\{ -\frac{1}{2\sigma_j^2}\left(x_j^2 + \frac{y_j^2}{q_j^2} \right) \right\} \tag{2.1}$$

with

$$\begin{aligned} y_j &= R\cos(\theta - \psi_j) \\ x_j &= R\sin(\theta - \psi_j) \end{aligned} \tag{2.2}$$

where (R, θ) are the polar coordinates in the sky plane and N is the number of Gaussians. The jth Gaussian has a total luminosity L_j, axis ratio given by $0 < q < 1$, dispersion along the major-axis σ_j, and position angle ψ_j. The algorithm then does a non-linear fit to the galaxy surface-brightness distribution.

Multi-component non-linear algorithms simultaneously fit the parametric surface-brightness distribution chosen for each galaxy component (*e.g.* the bulge, disc, bar, *etc*...) to the input data image. The majority of modern 2D photometric decomposition algorithms use the *Levenberg-Marquardt* (LM) fitting method which is based on the information given by a function gradient to assess the local structure associated with the space parameter to find the best solution efficiently. The LM algorithm is a specific minimisation method associated with multi-component non-linear fit. Despite LM being a powerful way to deal with non-linear fits, especially in terms of computer time, it suffers from some important drawbacks. Such drawbacks include it is ineffectiveness of jumping out of local minima.

There are however other algorithms one could use to overcome such difficulties. Simulated annealing (Neal 1996) looks for a best fit by generating random solutions in parameter space. The algorithm works in a similar fashion to Monte Carlo methods where if a set of parameters are considered as the 'better' solution, then the search continues from that solution. However, if the solution is considered to be worse then it is given a probabilistic opportunity to not be discarded. This is the premise of the Metropolis-Hastings algorithm.

Obtaining reliable parameter estimates is difficult, particularly in an automated fashion where resolution effects and the interplay between parameters cause disruption in the fitting process. Typically 20 to 30% of automatic fits are unrealistic so previous studies have used logical filters to address this problem (e.g. Allen et al. 2006, Simard et al. 2011; Meert et al 2015; Méndez-Abreu et al. 2016). Lange et al. 2016 listed five commonly occurring factors which lead to failures in the fitting process: i) Local minima trapping, ii) unrealistic solutions, iii) reversal of components, iv) indecisiveness to which model to use , and v) bad representation of final errors. Lange et al. used a grid of starting values combined with a convergence test to robustly obtain parameter values.

To circumvent these difficulties, we advocate embedding the galaxy morphology analysis into the broader context of inference and hypothesis testing with the use of a Bayesian Markov Chain

Monte Carlo (MCMC) technique. The above problems can then be solved in turn i.e. the exploration of parameter space can overcome runs that become trapped in local minima (i); initial priors can prevent unrealistic solutions (ii) and the reversal of components (iii); model comparison tests can help determine the most probable morphology (iv); and finally the outcome of the MCMC gives a proper description of the parameter uncertainties (v).

2.1.4 Chapter outline

This chapter introduces the use of an adaptive Bayesian MCMC algorithm and with emphasis on the methods, features and uses of such a technique. We demonstrate the feasibility of statistical inference based on obtaining galaxy morphology. A detailed exploration of the influence of the prior distribution, and explicit examples of model comparisons between single Sérsic and two-component bulge+disc models are also presented. The chapter is organised as follows. In Section 2.2, we describe the basic formalism of the inference methodology with an overview of Bayesian statistics and model generation. In Section 2.3 we give an overview of Bayesian statistics and a description of the likelihood and priors used for the new algorithm. We then describe the intricacies of the algorithm as well as how we achieve robust outcomes with convergence tests in Section 2.4. How we compare different model fits is discussed in Section 2.5. Finally, in Section 1.6 we discuss and summarise the algorithm presented.

2.2 Inference Methology

In this section, we describe the main attributes and implementation details of a new 2D galaxy photometric decomposition software code. For writing convenience the algorithm has been titled `PHI` (2D PHotometric decompositions using Bayesian Inference). The version of `PHI` used in this thesis is implemented using the *Interactive Data Visulisation* (IDL) software language. The flow chart for `PHI` is as follows:

1. `PHI` reads in either a list or single FITS file for the image(s) with their corresponding point spread function (PSF) and a map containing the sigma values for each pixel which corresponds to the error on the image.

2. The user can supply initial guesses to the parameters in the model as well as what type of prior for each parameter is desired. However, the need to input initial parameters is not essential for the algorithm to function correctly.

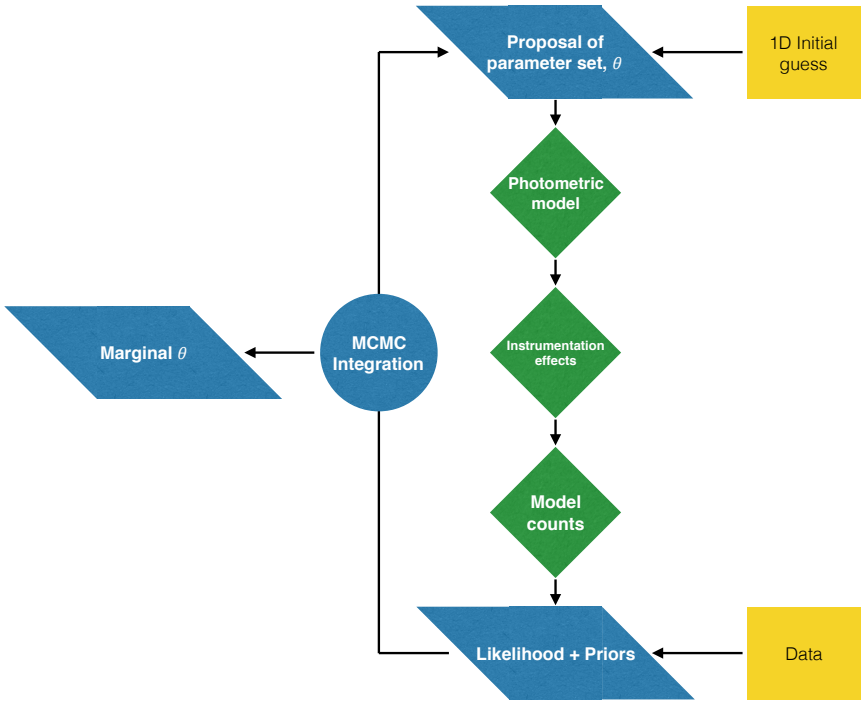


Figure 2.1: The control flow of the 2D photometric decomposition MCMC algorithm.

3. PHI then simulates the galaxy image with the user chosen models for each component. It then uses a fast fourier transfrom to convolve the model with the PSF image. The PSF can either be in a functional form, or user provided.

4. The likelihood and then the posterior probability are calculated for the current set of parameters. These values are fed into the MCMC engine of the code which is based in four phases of a MCMC algorithm. See Section 2.3 for further information.

5. Steps 3-4 are iterated as necessary until the full posterior has been mapped or the user defined iteration maximum is reached.

Figure 2.1 shows a summary flow chart of the above list. Each step will now be covered in detail.

2.2.1 2D-Photometric model functions and pixel sampling

As galaxies can be considered to be primarily two-component systems we will assume to first order that the observed surface brightness distribution of galaxies is composed by the sum of a bulge and a disc. For this chapter the definitions of the components are purely photometric, for example, the

bulge is described to be the excess of light over the inner extrapolation of an exponential disc. In this Thesis, elliptical galaxies are considered to be single component systems which have been commonly described in a similar fashion to bulges.

To represent the radial distribution of the stellar light *i.e* the surface brightness distribution, of each component, various mathematical functions are used. Over the years, observers have changed their opinion on what model function best describes which component. Today, there is a consensus on the function that best describes spheroidal objects such as the ordinary elliptical galaxy and the bulges of disc galaxies. This function is Sérsic's 1968 generalization of de Vaucouleurs' 1948; 1956 $R^{1/4}$ model to give the $R^{1/n}$ surface density profile. The Sérsic profile which describes how the projected surface-intensity I varies with the projected radius R has the form,

$$I(R) = I_e \exp\Big\{-b_n\Big[\Big(\frac{R}{R_e}\Big)^{1/n} - 1\Big]\Big\}, \tag{2.3}$$

where I_e is the intensity at the effective radius R_e that encloses half of the total light from the model (Ciotti (1991); Caon et al. (1993)). b_n is defined in terms of the third parameter n, the Sérsic index, which describes the concentration of the light profile. When $n = 1$ the model follows an exponential surface-intensity profile and $n = 4$ reproduces the de Vaucouleurs' model; thus the Sérsic profile can describe the main body of observed spheroidal objects. The term b_n is estimated to be $b_n \approx 1.9992n - 0.3271$ within the range $0.5 < n < 10$ (Capaccioli 1989). The exact value of b_n can be obtained by solving the complete gamma function $\Gamma(2n) = 2\gamma(2n, b_n)$, where $\gamma(2n, b_n)$ is the incomplete gamma function (Ciotti, 1991). A detailed review of Sérsic's profile plus associated quantities has been provided by Graham & Driver (2005).

Galaxies in the Universe can be broadly distinguished into two groups; spheroids and discs. It was discussed above that spheroidal objects can be described by the Sérsic profile, with observed objects spanning a range of concentrations and sizes. Galaxies with large-scale stellar discs are termed 'disc galaxies' and commonly have centrally located stellar distributions (*i.e* the bulge) that appears as an excess from the relative inward extrapolation of the outer exponential disc light. The exponential extent of the radial distribution of the starlight from the disc component of disc galaxies has been known for sometime (de Vaucouleurs, 1956; Freeman, 1970), with the intensity I changing with R in the form

$$I(R) = I_0 \exp\Big\{-\frac{R}{h}\Big\}, \tag{2.4}$$

where I_0 is the central intensity and h is the e-folding disc scale length. As mentioned above, a value for $n = 1$ for the Sérsic profile will also achieve an exponential model (Graham & Driver, 2005).

For this work the components of galaxies are characterised by elliptical and concentric isophotes with constant (and likely different) ellipticity $\epsilon = 1 - q$ (where q is the ratio between the semi-minor and semi-major axis of the ellipse) and position angle $\theta_{\rm PA}$ in degrees counter-clockwise from the vertical axis of the image. The image function deals with a converted ellipticity of the form $q = b/a (= 1 - \epsilon)$ and the position angle as the angle relative to the image x-axis PA ($= \theta_{\rm PA} + 90\,{\rm deg}$) The isophotes are centred on (x_0, y_0) with the projected radius given by

$$r = \left(x_p^2 + \frac{y_p^2}{q^2} \right) \tag{2.5}$$

where x_p and y_p are coordinates in the reference frame centred on the image-function centre (x_0, y_0) and rotated to its position angle:

$$\begin{aligned} x_p &= (x - x_0)\cos({\rm PA}) + (y - y_0)\sin({\rm PA}) \\ y_p &= -(x - x_0)\sin({\rm PA}) + (y - y_0)\cos({\rm PA}) \end{aligned} \tag{2.6}$$

This projected radius is then used to compute the surface brightness distributions as discussed above.

Most published references state the need to oversample the central pixels of the image. This is due to the sharp gradient in flux between adjacent pixels in the centre. Oversampling a fixed region or the entire image can be highly inefficient. GALFIT oversample pixels based on their distance from the centre of the image with the exception of Nuker profile Peng et al. (2002) while PROFIT uses the gradient along the minor axis of the profile and utilises a multilayered oversampling approach Robotham et al. (2017). Figure 2.2 and Fig.2.3 show how increasing the Sérsic index increases the gradient of the profile in the central regions. `PHI` adapts the oversampling region by calculating the Euclidean norm of the gradient vector at every pixel of the image using the adjacent pixels. Where the gradient is over a tolerance specified by the user the pixel is subsequently flagged. Flagged pixels are divided into smaller grids (10×10 sub-pixel grids) with each sub-pixel becoming redefined.

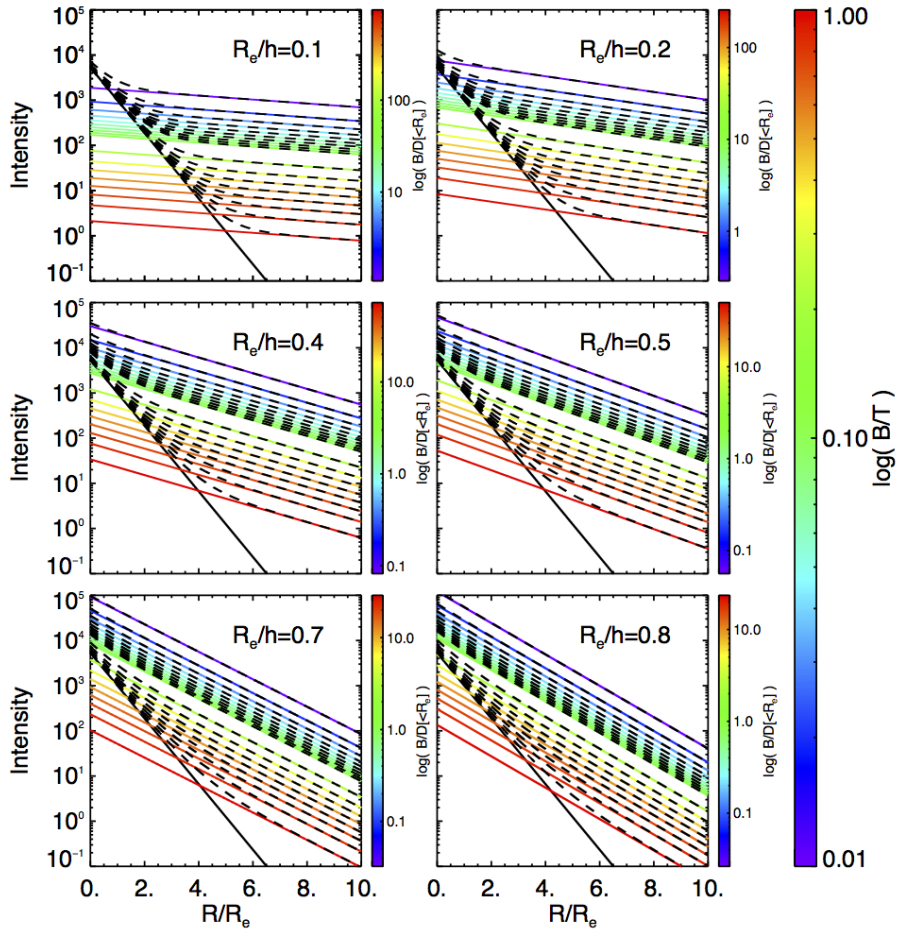


Figure 2.2: A range of two-component models with the intensity along the y-axis and the radius normalised with the effective radius of the Sérsic profile on the x-axis. The black solid line is a Sérsic profile with $n = 1$. Each frame has a different value of the ratio R_e/h with exponential component designated with the coloured lines. The colour of the line represents both the B/T shown in the larger colour bar on the right and the B/D within one effective radius shown by the individual colour bars to the right of each frame.

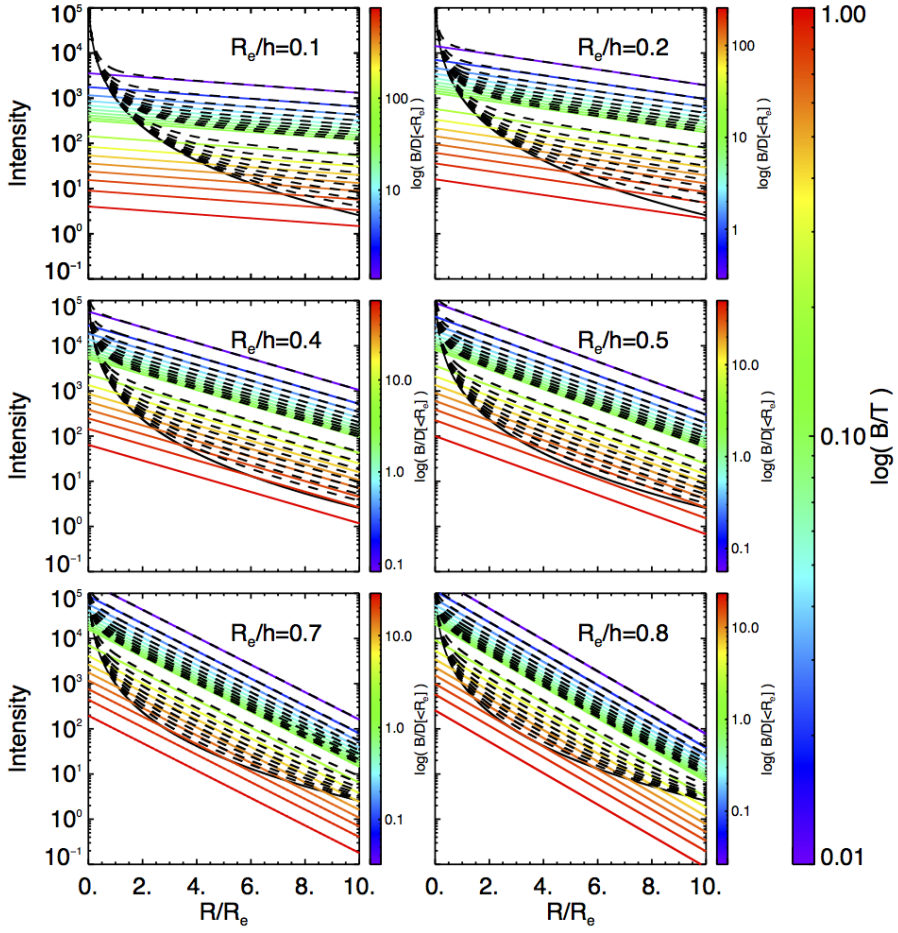


Figure 2.3: A range of two-component orientations. See the caption to Fig. 2.2. The solid line shows a Sérsic profile with $n = 4$.

2.2.2 The sky background

The sky background is a flat plane along the x- and y- image dimensions. PHI has the capability of including the sky background value in the fitting process. Méndez-Abreu et al. (2008a) shown that the most significant contributors to parameter miscalculation where systematic errors in the sky background and PSF FWHM. Improper sky subtraction was found to affect the the disc surface brightness profile in a two-component model. The circumvent these systematic errors Méndez-Abreu et al. (2008a) advised a careful sky subtraction pre-fit.

2.2.3 Convolution with the PSF

A careful analysis of the point spread function (PSF) is needed to perform robust photometric decompositions. This has been shown in many previous studies. Méndez-Abreu et al. (2008) found that errors of 2% in the PSF FWHM can produce variations of up to 10% in the R_e and n bulge parameters. (Gadotti, 2009) showed that to reliably retrieve the structural properties of bulges the effective radius, R_e has to larger than 80% of the PSF full width half maximum. Our algorithm uses a Fast Fourier Transform (FFT) to convolve the PSF with the model image. There are alternative convolution techniques, however, we find the FFT convolution is fastest and thus used in the remainder of this paper.

2.3 Bayesian Markov Chain Monte Carlo

Bayesian statistics puts probabilities at the forefront. The evidence about the true state of the world is expressed in terms of Bayesian probabilities which can be thought of as a degree of knowledge. Bayesian methods use the combination of expert scientific information of a model characterised by its parameter vector θ along with data, D, to obtain probabilities; this is described in Bayes Theorem,

$$p(\theta|D) = \frac{L(D|\theta)p(\theta)}{\int L(D|\theta)p(\theta)d\theta}, \tag{2.7}$$

where $p(\theta|D)$ is the posterior probability distribution of having a set of parameters θ given the data D; $L(D|\theta)$ is the likelihood function or the probability of the data given θ; and $p(\theta)$ is the prior probability of the parameter vector θ. We are free to assign any form of $p(\theta)$ that we feel best represents our understanding of the model. The prior distribution expresses our knowledge and prejudices about the relative likelihoods of the models or of specific parameter values of θ. The denominator of Equation 2.7 is just a normalisation factor to ensure the probability is of unity when

summed over all possible models, it is sometimes referred to as the probability of the data.

Bayesian methods are restricted by the need to perform integrations analytically. For a high-dimensional parameter space such as the models used to describe the surface brightness distribution of galaxies, the characterisation of the posterior distribution would become infinitely difficult. In recent decades, approximate Bayesian analysis has been performed by using numerical integration or sampling-based estimation methods, *i.e*, Monte Carlo methods. By generating repeated states by a first-order Markovian process, a Markov-chain Monte Carlo (MCMC) will asymptotically converge to the posterior distribution. Such sampling provides probabilities relating to θ.

The goal of a MCMC in Bayesian inference is to maximize the unnormalized joint posterior distribution and collect samples of the target distributions, which are marginal posterior distributions, later to be used for inference. In the simple case where we might have a model with two or three parameters, one can analytically calculate the posterior probability by forming a grid of parameter values to explore all the possible values. This is a common method in Astronomy, however, for a high dimensional parameter space such as fitting a Sérsic and an exponential surface brightness profile to a spiral galaxy image (9 free parameters), this would take a significant amount of time. We can use MCMC to explore this high-dimensional parameter space more efficiently than ever before due to improvements in the computational method.

2.3.1 Likelihood & priors

If we assume that a large number of photons are being detected in the CDD then the measurement noise can be considered to be Gaussian with a mean zero and covariance matrix C. It then follows that the likelihood can also be written as the Gaussian probability density function (PDF) for the difference between measurements and observations (Tamminen, 1999):

$$p(D|\theta) = \frac{1}{(2\pi)^{N/2}(\det C)^{1/2}} \times \exp\Big(-\frac{1}{2}(D - f(x;\theta))^T C^{-1}(D - f(x;\theta))\Big) \tag{2.8}$$

where $f(x;\theta)$ is the model function which consists of known quantities x (*i.e* constants, control variables etc.) and the unknown parameters θ and N is the total number of pixels. If we assume that the measurement errors ($\epsilon_i = d_i - f(x_i;\theta)$) are normally distributed and independent such that $\epsilon_i \sim N(0,\sigma^2)$, the likelihood for a certain measurement gets the form

$$p(d_i|\theta) = \frac{1}{(2\pi\sigma_i^2)^{1/2}} \exp\Big(-\frac{1}{2\sigma_i^2}(d_i - f(x_i;\theta))^2\Big). \tag{2.9}$$

Since the error terms are considered to be independent, the combined likelihood of all the measurements can be written as a product

$$p(D|\theta) = L(D|\theta) = \prod_{i=1}^{N} \frac{1}{(2\pi\sigma_i^2)^{1/2}} \exp\Big(-\frac{1}{2}\chi^2\Big). \tag{2.10}$$

where χ^2 takes the form

$$\chi^2 = \sum_{i=1}^{N} \frac{(d_i - f(x_i;\theta))^2}{\sigma_i^2}. \tag{2.11}$$

For practical reasons the likelihood is calculated in its logarithmic form so as to increase the numerical accuracy. The final form of the likelihood used for the remainder of this paper is

$$\ln L(D|\theta) = -\frac{\chi^2}{2} - \sum_{i=1}^{N} \ln\sigma_i - \frac{N}{2}\ln 2\pi. \tag{2.12}$$

As mentioned above, the prior distribution describes our a priori (previous) knowledge about the model, or more specifically θ. Where we know some values to be more probable than others, a carefully selected prior distribution can emphasise this knowledge. However, biases may arise if the prior distribution is *informative e.g.* a Gaussian PDF with a narrow width. For the purpose of galaxy image analysis, where the parameter space is known to have many local minima, it is inadvisable to use such an informative prior.

If we do not have any prior knowledge or if we do not want to impart a bias into the fitting process then an *uninformative* prior can be used. The best example of an uninformative prior distribution is a uniform distribution:

$$p(\theta) = U(a,b) = \begin{cases} \frac{1}{b-a} & \text{for } \theta \in [a,b] \\ 0 & \text{otherwise} \end{cases}. \tag{2.13}$$

where a and b are parameter limits set by the user. Thus it can be considered that $p(\theta)$ is constant within the limits a and b, and consequently returning to a maximum likelihood estimation in a practical sense.

In the case of photometric decompositions of galaxy images there are some physical motivated limits on the parameters that we should include in their priors. For example, negative or very large

Table 2.1: Input parameters and priors used in the MCMC code.

Individual parameters			
Parameter	Symbol	Prior	Range
Effective intensity	I_e	Uniform in $\log(I_e)$	$\log(I_e/counts) \in U[0.01, 10]$
Effective radius	R_e	Uniform in $\log(R_e)$	$\log(R_e/pixels) \in U[0.01, 10]$
Sérsic index	n	Uniform in n	$n \in U[0.4, 8]$
Central intensity	I_0	Uniform in $\log(I_0)$	$\log(I_0/counts) \in U[0.01, 10]$
Scale length	h	Uniform in $\log(h)$	$\log(h/pixels) \in U[0.01, 10]$
Axial ratio	q	Uniform in q	$q \in U[0.2, 1]$
Position angle	PA	Uniform in PA	PA/degrees $\in U[-360, 360]$
Central coordinates	x_0, y_0	Uniform in x_0&y_0	x_0&$y_0 \in U[0, S_{image}]$
Combined parameters			
Effective radius / scale length	R_e/h	Uniform in R_e/h	$R_e/h \in U[0.05, 1.678]$
Bulge-to-total ratio	B/T	Uniform in B/T	$B/T \in U[0.01, 1]$
Bulge-to-total ratio ($< R_e$)	$B/T(< R_e)$	Uniform in $B/T(< R_e)$	$B/T(< R_e) \in U[1, -]$
# of crossing points	N_x	Delta funtion (see text)	

values for the radius are unphysical. Physical values for the Sérsic index, n, are within the range $0.5 < n < 8$ as larger values will produce non-physical concentrations of stellar light in the centres of galaxies. A clear cut-off point can be put in place to prevent non-physical parameter values *e.g* negative radii. Tables 2.1 list the prior functions used for each parameter of each component.

Figures 2.2 and 2.3 present the effect of varying each of the different model parameters on the intensity profiles of two-component galaxies. We define a two-component system as a combination between an inner Sérsic with an outer exponential surface brightness profile. For fitting, we wish to consider all possible combinations where the inner regions are dominated by the inner (bulge) component and vice versa for the outer (disc) regions. While this sounds straightforward, for an inner Sérsic profile with $n > 1$ and negatively sloped exponential disc profile, at some (large) radius the inner component will again dominate over the outer. The definition of where a galaxy's outer edge is can be difficult to quantify. However, for photometric decompositions we work only with the current image we are fitting, thus, so long as the outer profile is dominant before the object itself vanishes into the background noise our canonical definition of the galaxy is met. In the case where the object is fit with a single component all the above priors relating to the structural parameters apply i.e. keeping all sizes and intensities positive along with the range in Sérsic index values and geometric properties.

The reversal of the bulge and disc in two-component fittings is a problem we want to avoid. This is where the initially assigned bulge component migrates to fit the disc and the disc to the bulge (Allen et al. 2006). Figure 2.2 and 2.3 show this in practice when the B/T is very low or when the

B/D within one effective radius is greater than 1. Here we start with the prior assumption that the inner profile will represent the bulge and the outer the disc. The algorithm implements this prior knowledge in three ways: i) preventing the disc effective radius ($R_{disc} = 1.678h$) becoming smaller than the effective radius of the bulge, ii) making sure the bulge-to-disc ratio within one effective bulge radius is dominant (i.e. $B/D \geq 1$); and iii) preventing the light profile of the bulge becoming dominant in the outer edges of the image. We can see an example of this in Figure 2.3 when both B/T and R_e/h of the bulge becomes more dominant at larger R. This final prior is implemented using a Newton-Raphson algorithm to determine crossing points in the bulge and the disc light 1D profiles with the following prior applied:

$$p(\theta) = \delta(N_x) = \begin{cases} 1 & \text{for } N_x = 1 \\ 0 & \text{otherwise} \end{cases} \tag{2.14}$$

where $\delta(N_x)$ is a delta function on the number of crossing points N_x in the 1D photometric profile. The Newton-Raphson algorithm only implemented until the total model galaxy light function falls below the mean of the sky background noise; after this a second crossing point can occur (and is inevitable for galaxies with $n > 1$).

For the remainder of this thesis these combined priors are applied to all synthetic and real galaxy fit. However, for future studies the code is organised so that priors can be easily modified.

2.4 The MCMC engine

The MCMC algorithm used in PHI consists of three levels that aim to achieve an efficient convergence and accurately estimate the posterior distribution. Fig. 2.4 shows a typical run of the entire algorithm with different colours depicting the three levels and transitions between them. In the following subsections we will address the intricacies of each level individually.

2.4.1 Level one: Blocked Adaptive Metropolis

PHI begins with a variation on the Adaptive-Metropolis-within-Gibbs which was introduced in Roberts & Rosenthal (2009). The purpose on this level is to obtain an estimate on the scale of each parameter or coordinate in the Markov chain. By knowing how each coordinate scales PHI can efficiently sample from the parameter space. The framework of the Metropolis algorithm (Metropolis-within-Gibbs is fundamentally the same) is as follows. Given a current value in the Markov chain, X_i (*e.g.*

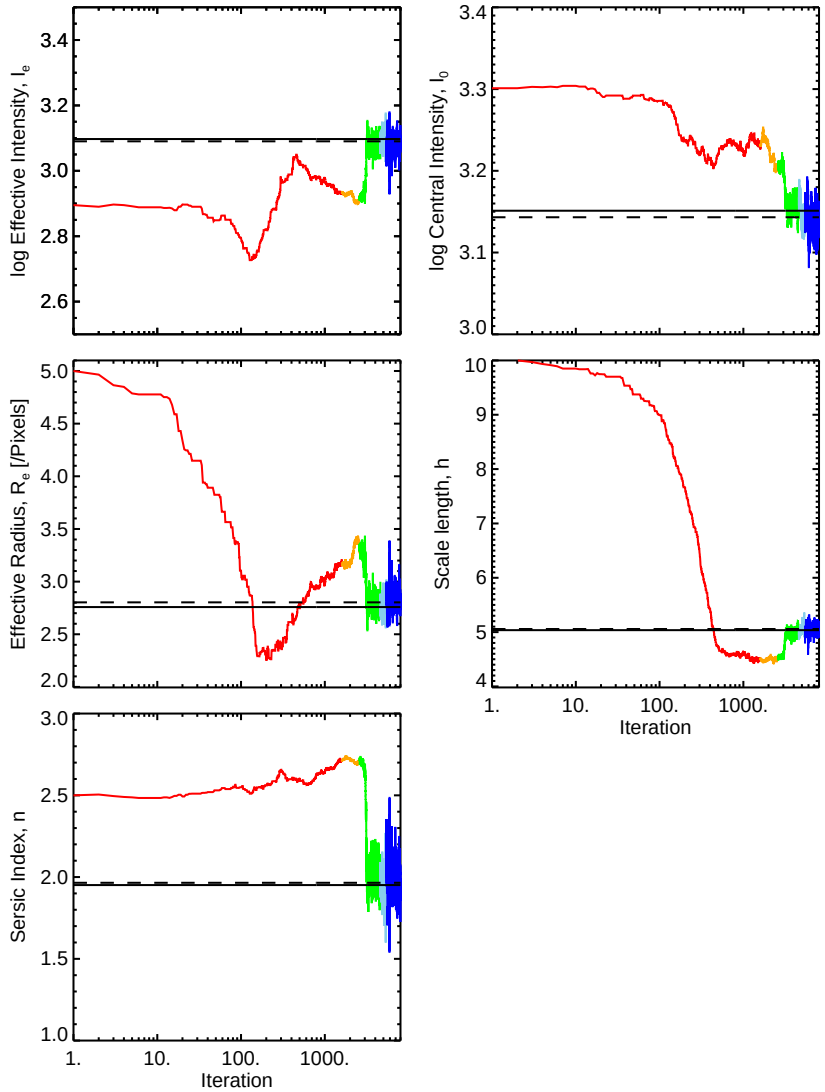


Figure 2.4: Trace plots for the structural parameters of a two-component model. Each plot represents one parameter. The left column shows the Sérsic profile parameters; the effective intensity, I_e, the effective radius, R_e, and the Sérsic index, n. The right column shows the exponential profile structural parameters; the central intensity, I_0 and the scale length h. The red, orange, green and blue corresponds to the 1^{st}, the transition, the 2^{nd} and the final level of the algorithm. The light blue segment represents the burn-in which is discarded with only the final dark blue segment taken as a sample.

the parameter values at some stage θ_i) a new value or set of values Y is proposed. If i denotes the i^{th} Markov chain then j denotes the j^{th} parameter or group of parameters of interest. The proposed chain coordinates are drawn from $Y_j \sim N(X_{i,j}, \sigma^2)$. Y is either accepted as a valid move so the next starting location is $X_{i+1} = Y$ or it is rejected and $X_{i+1} = Xi$. Formally the criteria for acceptance is,

$$\begin{cases} X_{i+1} = Y & \text{if } U < min[1, \pi(y)/\pi(x_i)] \\ X_{i+1} = X_i & \text{if } U \geq min[1, \pi(y)/\pi(x_i)] \end{cases}. \tag{2.15}$$

where U is a uniformly chosen random number $U \sim U(0,1)$ and $\pi(\centerdot)$ is the target distribution. Practically speaking $\pi(\centerdot)$ is the numerator in Equation 2.7 *i.e.,* the combination of the likelihood and the prior distribution for a given state. This process is repeated for every coordinate j, sequentially. As one of the challenges in MCMC sampling is achieving an exploration of the parameter space in the most efficient way, by grouping the most correlated parameters the chains can quickly traverse the stationary distribution. The structural parameters were found to correlate strongly with each other (i.e. I_e, R_e, n, I_0, and h). Geometric parameters show moderate correlation with each other while having insignificant impact on the structural parameters. For the remainder of this thesis we have two groups of blocked parameters: the structural and the geometric parameters.

The initial conditions X_0 for the chain can vary, either they can be user given, estimated from a one-dimensional fit or chosen randomly. The problem with the initial conditions is less of a problem with the MCMC but more to do with the model. However, the MCMC can solve problems of the initial conditions by efficiently and effectively exploring the parameter space. It is the goal of adaptive algorithms to improve the Markov chain in order to accomplish this. By using information from past iterations we can significantly improve efficiency in practice (e.g. Haario et al., 2001; Giordani & Kohn, 2010; Roberts & Rosenthal, 2009; Vihola, 2012). For each coordinate j of a Markov chain the optimal acceptance rate is 0.44 for a one-dimensional Markov chain and 0.23 for dimensions greater than one (Roberts & Rosenthal, 2001). As stated above, a proposed chain Y_j is drawn from a normal distribution with σ_j. The σ_j can be thought of as the size of the step the algorithm makes when choosing the proposed values. If σ_j is too large then we will see a drop in the acceptance rate as we will be drawing from a region of parameter space with low probability and hence it will be less likely to be accepted. Vice-versa from a σ_j which is to small, now we will be accepting values almost at every iteration. So we are presented with a compromise between being able to jump from one region of parameter space to another quickly and being able to explore in detail the target distribution. It is undesirable to have the extremes of both scenarios as the chances of being

stuck in local minima increases. So by adapting σ_j until the desired acceptance rate is achieved is by far the most efficient course of action. Roberts & Rosenthal (2007) showed that for an adaptive algorithm to be *ergodic* the total variation T between the distribution of the adaptive algorithm at time n and the target distribution should tend to zero *i.e.,* $\lim_{n\to\infty} T = 0$. This means for a Markov chain to be considered 'valid' it will eventually need to converge to the stationary target distribution $\pi(\cdot)$. Theorem 1 from Roberts & Rosenthal (2007) gave some conditions for a Markov chain to be ergodic. For example the amount of adaptation should diminish i.e. $\lim_{n\to\infty} A_n = 0$ where $A_n = \sup ||P_{\Gamma_{n+1}}(x,\cdot) - P_{\Gamma_n}(x,\cdot)||$ and Γ is the random variable which updates how the Markov chain steps in parameter space. To accomplish diminishing adaptation we initially allow an update to σ_n when the acceptance rate is higher or lower than 0.23 (or 0.44 for Gibbs-like movements) for n_{step} iterations (the default $n_{step} = 100$ but can be modified by the user). If the calculated acceptance rate is higher than what is desired then 5% of σ_n is added to σ_n and 5% is subtracted if it is lower. Once the acceptance rate falls within 0.15 and 0.32 for blocked parameters (or 0.28 and 0.6 for Gibbs-like) adaptation will occur every $n_{step} = 2n_{step}$ iterations taking the past $2n_{step}$ for the acceptance rate calculation. As before the calculated acceptance rate for each coordinate must fall within 0.15 and 0.32. This step is repeated again with $2n_{step}$ iterations being used in the calculation. With every coordinate having acceptance rates within 0.15 and 0.32, adaptation is stopped and the σ_n is saved.

It is important that the Markov chain is close to the target distribution, so the chain follows from before but without adaptation to ensure this. Again from a current value of the Markov chain X_n, Y is proposed by generating a set of randomly generated parameter groups $Y_j \sim N(X_{n,j}, \sigma_j^2)$, with$\sigma_j$ staying fixed. For Level one to finish the gradient of the chain must tend to zero, i.e. the chain is converging to the target distribution. Once this has been satisfied the algorithm can move onto Level two.

2.4.2 Level two: Adaptive Metropolis

The aim of this level is to obtain a similar covariance structure for the proposal distribution (Y) to that of the target distribution, which leads to greater success rates for the proposal distribution (Haario et al., 2001; Roberts & Rosenthal, 2009). Y is drawn in a similar way as before i.e. $Y \sim N(X_n, c\Sigma_n)$, again where X_n is the current state of the chain, and the same accept/reject Metropolis rule is used. $c\Sigma_n$ is the covariance matrix of all the previously generated values of the chain since the adaptation of level one finished multiplied by a constant $c = 2.382^2/d$ (see Haario et al. (2001) and Roberts & Rosenthal (2009)) where d is the number of fit parameters or parameter dimensions. The multiple $2.382^2/d$ yields an optimal acceptance rate, if multiplied by the target covariance matrix

to obtain the proposed covariance matrix (Roberts & Rosenthal, 2009). The proposed distribution originates from a multivariate normal distribution (MVD) as to generate a set of possibly correlated real-valued random variables clustered around some mean value. As the contours of a MVD so are linear transformations of hyperspheres centred at the mean they can be considered as ellipsoids. The eigenvectors of the covariance matrix give the directions of the principal axes of these ellipsoids with the corresponding eigenvalues giving the squared relative lengths of the principal axes. If we consider Σ_n to be a real symmetric matrix, the eigenvalues are real and the eigenvectors can be chosen such that they orthogonal to each other. Thus a real symmetric matrix and hence Σ_n can be decomposed as

$$\Sigma = U\Lambda U^T = U\Lambda^{1/2}(U\Lambda^{1/2})^T, \tag{2.16}$$

where U is the unit eigenvectors and Λ is a diagonal matrix of the eigenvalues, we then have

$$Y \sim N(X_n, c\Sigma_n) \Longleftrightarrow Y \sim X_n + U\Lambda^{1/2}N(0, I), \tag{2.17}$$

as $N(X_n, c\Sigma_n)$ is effectively a normal distribution of the unit vector I, i.e. $N(0, I)$, scaled by the diagonal matrix Λ containing the eigenvalues, rotated by the eigenvectors U and then translated by X_n. Assuming again that Σ_n is a positive semi-definite matrix we can use a specific eigen-decomposition called the Cholesky decomposition, $\Sigma_n = LL^T$, where L is the lower triangle and L^T is just the transpose of L. Comparing this with equation 2.16 it can be said that the Cholesky decomposition of Σ_n will give us $U\Lambda^{1/2}$, which can then be substituted into equation 2.17. Using the Cholesky decomposition will decrease computational time in relation to the alternative *i.e.,* calculating the eigenvectors and eigenvalues.

To establish if further adaptation will improve the chain or not the algorithm will test that the covariance structure of the target distribution has been correctly identified. This can be determined directly from the past iterations of the chain. Every N_{L2} (user input) iterations the mean squared difference between each successive iteration $\langle (X_{n-1} - X_n)^2 \rangle$ for each parameter coordinate is calculated, and after $5N_{L2}$ a linear model is fit. If the gradient of the mean squared differences appears to have an increasing or decreasing gradient then the algorithm continues to adapt;if the gradient is close to zero then adaptation stops and the algorithm moves to the final level.

2.4.3 Final level & chain convergence

The final level of the PHI applies a symmetric random walk Metropolis algorithm drawing the proposed values from $Y \sim N(X_n, \Sigma_{L2})$. Σ_{L2} is the last covariance matrix calculated before adaptation stopped in level two multiplied by the constant c. The parameter distributions of this final stage are an estimate of the target distribution and thus will be taken for the final analysis.

There are multiple ways to test for convergence of the Markov chains to the target distribution. One way is to let one chain test the parameter space for a large number of iterations. A Geweke diagnostic (Geweke, 1992) can then determine if the Markov chain has converged to the target distribution.

The alternative way of obtaining a converged sample distribution is to allow multiple chains to run simultaneously, this is the default method for PHI. The starting values for the chains are drawn from a uniformly random number with the limits being the maximum and minimum values accepted in Level two. A Geweke diagnostic can test for convergence in single chains but to test if multiple chains have converged we use a Gelman-Rubin diagnostic. Once the Markov chains have converged, the chains are combined to form the final sample distribution that will be used in the analysis stage.

In a typical run, PHI requires between 10^4 to $10^{4.5}$ iterations for three simultaneously running chains to converge. The median total generation time for a 250×250 pixel image is $t_{total} = 0.029s$ for a single Sérsic model and $t_{total} = 0.041s$ for a Sérsic + exponential model. The wall clock time for a complete run on a 2.5 GHz Intel core i5 CPU is $\sim$ 10 minutes and $\sim$ 20 minutes for a single Sérsic and a Sérsic + exponential model, respectively. Run times are similar for real and mock galaxies.

2.4.4 Posterior probabilities

A reliable error estimate is essential when quantifying trends in the derived properties of galaxies. The Bayesian MCMC approach samples from the target distribution,, which is later used to derive reliable error estimates for all the model parameters describing the data. Visually the full multivariate distribution can be used to derive errors, however, the medians and percentiles can be used for a quantitative measure of the posterior distribution. Parametric surface brightness models invariably result in parameter covariance which can then be exaggerated by instrumental and selection effects. Our approach explicitly incorporates parameter covariance, noise sources and other instrumental effects including signal-to-noise, and PSF convolution to result in a reliable inference as discussed

above.

Figure 2.5 shows an example of a marginalised posterior produced by the MCMC method. The model used to describe the data includes all model parameters in a two-component system i.e. an inner Sérsic with an outer exponential component. Figure 2.5 shows all the structural parameters: the effective intensity I_e, the effective radius and the Sérsic index for the inner component and the central intensity I_0 and the scale length h for the outer exponential component. The marginal distribution for each model parameter is shown on the diagonal, and the joint marginal distributions of parameter pairs are shown in the off-diagonals with contour colours corresponding to the two-dimensional probability density. The sample used to build the posterior distributions ran three simultaneous chains to ensure convergence. The location of the true value for this synthetic galaxy (see Section 3.3 for the synthetic galaxy generation) is indicated by the solid black line along with the median of the posterior distributions shown by the dashed line.

For this resolved galaxy, it is clear that there is a strong covariance between parameters within the individual components i.e. I_e, R_e, and n for the Sérsic profile and I_0 and h for the exponential component. The posteriors are close to the true values with the medians only ever having a fractional error of 2% in relation to the true values.

The Bayesian based procedure presented in this paper explicitly incorporates the parameter covariance present in the models. Furthermore, it enables us to utilise the entire posterior distribution for a galaxy in a hierarchical model to readily test hypotheses about the galaxy population. This will be demonstrated in a following chapter. For this current study however, we can use the posterior distributions to better understand the real causes of the parameter covariances and biases that make the posterior medians drift from the true values. For an analysis it is sometimes impractical to use the full posterior distribution so to represent the best fitting model we use the median of the final sample distribution and use the first and third percentiles to describe the errors.

2.5 Bayesian model comparison

As previously stated, our approach was to set up a simple model that uses a small number of parameters. Once the models have successfully been fitted to the data we can increase the complexity of the model i.e. adding an extra component. In Gelman et al. 2003 the statistical theory of hypothesis testing involves methods that aim to check whether an improvement of a fit is statistically significant. Model comparison methods sometimes use the Bayes factor. Suppose we have two competing models for the data, model M_0 has parameters θ_0 and the competing model M_1 has the parameters

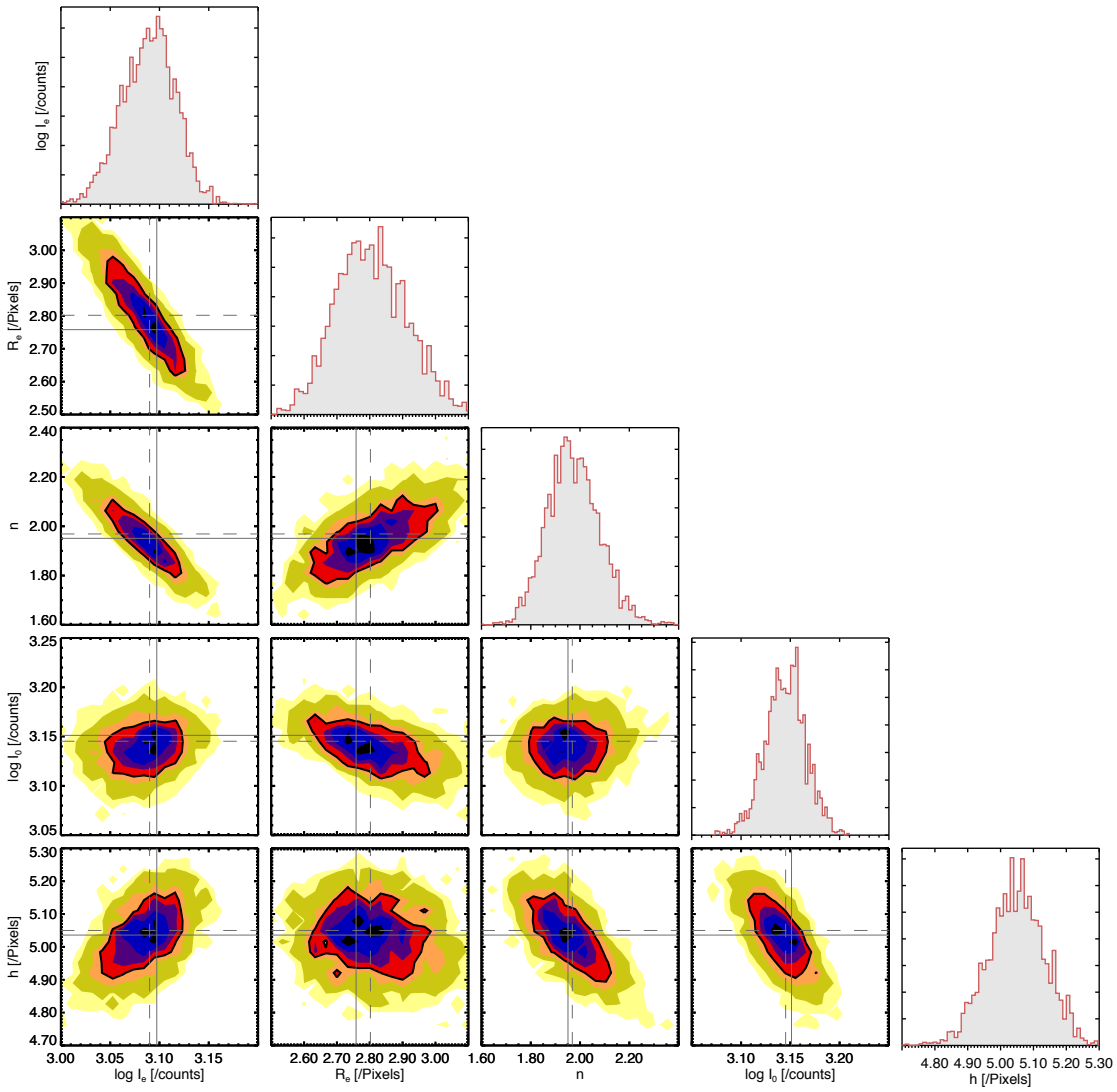


Figure 2.5: Posterior marginals for a synthetic galaxy. The full marginal distribution for each of the structural parameters is shown on the diagonal. Joint marginal pairs of parameters are shown on the off-diagonals. The seven color contours represent the 10, 30, 50, 68, 80, 95, and 99% confidence levels and the black solid line signifies 68% confidence level. The solid grey line shows the true value of this synthetic galaxy and the dashed line represents the median of the posterior distribution. There is a strong covariance between all the Sérsic profile parameters as well as the exponential profile parameters. Although, evidence for a weaker covariance between the two components also exists.

θ_1. Using Bayes theorem we calculate the posterior odds or the ratio of the posterior probabilities as

$$\frac{p(\theta_1|D)}{p(\theta_0|D)} = \frac{p(\theta_1)}{p(\theta_0)} \times \text{Bayes factor}(M_1, M_0) \tag{2.18}$$

where

$$\text{Bayes factor}(M_0, M_1) = \frac{p(D|M_0)}{p(D|M_1)} = \frac{\int p(\theta_0|M_0)p(D|\theta_0, M_0)d\theta_0}{\int p(\theta_1|M_1)p(D|\theta_1, M_1)d\theta_1}. \tag{2.19}$$

Bayes factors are most appropriate for a Bayesian methodology but are notoriously difficult to compute due to need to integrate over the entire parameter space. Bayes factors are also sensitive to the prior specification. However if performed the Bayes factor can assign probabilities to competing models providing us with a quantitative galaxy classification method.

To overcome some of the difficulties in calculating the Bayes factor we explore an alternative model comparison method. Schwarz (1978) proposed the *Bayesian information criterion* (BIC) as providing a large sample approximation to the Bayes factor. The BIC approch compares the maximum of the likelihood $L(D|\theta)$ of each model modified by the number of degrees of freedom m. Specifically

$$\text{BIC} = -2\ln(L(D|\theta)) + m\ln(N), \tag{2.20}$$

where N is the number of data points used in the fit. The application of the BIC aims to find the optimal number of degrees of freedom that explains the data. The BIC has also been argued to penalise over-fitting more than other model comparison methods (Raftery, 1995), thus the method will prefer lower degrees of freedom. We can use the difference between the BIC of M_0 and that of M_1 to check whether one model is significantly better than another, given the different number of degrees of freedom:

$$\Delta\text{BIC}_{10} = -2\log\left(\frac{p(D|\theta_0, M_0)}{p(D|\theta_1, M_1)}\right) - (m_1 - m_0)\log(N), \tag{2.21}$$

where model 1 will be chosen only when the ΔBIC_{10} statistic is large. We approximate ΔBIC_{10} by fitting both models to the object and taking the posterior medians for the fitted parameters θ_i

rather than the maximum likelihood. For the remainder of this thesis we will use the BIC to give an indication on what model best describes the structure of a galaxy. In reality galaxies have a variety of complex features (e.g. spiral arms, star forming clumps, *ect* ...) which could effect the BIC value. Also the BIC is only an approximation of the Bayes factor (and practically is not Bayesian) so further tests need to be preformed in order to assess it's predictive power in the field of galaxy decompositions (See Chapter 3).

2.6 Summary

In this chapter we have presented a new fitting algorithm (`PHI`) to perform 2D photometric decompositions of galaxy images from a Bayesian perspective. The algorithm is implemented to run an adaptive MCMC for a prescribed amount of time, diagnose when adaptation is sufficient, and then run a conventional MCMC with an estimated covariance matrix to better explore the parameter space. Convergence diagnostics are also used to ensure a robust estimation of the target posterior probability distribution.

Our approach offers a number of significant advantages for estimating surface brightness profile parameters. Algorithms that use standard downhill optimisation techniques can have five commonly occurring factors which lead to failings in the fitting process: i) Local minima trapping, ii) unrealistic solutions, iii) reversal of components, iv) indecisiveness to which model to use , and v) bad representation of the final errors. `PHI` addresses each problem as follows:

I. `PHI` incorporates a triple layer approach. The first layer uses a blocked adaptive Metropolis algorithm to obtain an estimate of the scale for each parameter in the chain. The second layer uses an adaptive Metropolis algorithm with the purpose of estimating the target covariance matrix. We assume the proposed distribution can be described as multivariate normal distribution. The final level uses this calculated covariance matrix to quickly and effectively explore the parameter space reducing the chances of a local minima trap.

II. We have implemented a number of priors that aim to allow the parameters to stay realistic and physical (i.e. positive in the case of the dimensions and intensities). These priors are better understood as boundary regions similar to the filtering process used in past work to remove non-physical parameter outcomes.

III. To prevent the reversal of components (i.e. the desired inner component profile switching to fit the outer and vice versa) we use a combination of priors. A Newton-Raphson algorithm

determines the crossing points in the total light profile and calculates the dominant component in the centre regions. This prior combination specifies that the bulges of galaxies are better modelled by a Sérsic profile and the discs are described by an exponential profile.

IV. Finally, `PHI` gives the full posterior probability distribution for a set of model parameters. This is a powerful description of the model uncertainties that can be used in further analysis of galaxy structures.

For future studies a full Bayesian analysis of galaxy morphologies is essential in unlocking the remaining unanswered questions about galaxy structures. With the addition of `PHI` into the array of 2D photometric decomposition toolbox we hope to improve our understanding of galaxy properties.

3

A Bayesian Approach to 2D Photometric Decompositions: Applications to synthetic & real galaxies

3.1 Introduction

In the previous chapter we introduced the new 2D photometric decompositions code, `PHI`. As previously stated, a maximum-likelihood analysis can erroneously imply correlations owing to the complexity of the parameter space when dealing with multi-component fits of galaxy images. The mechanics of the Markov Chain Monte Carlo (MCMC; more specifically the Metropolis-Hastings algorithm) allows for a more rigorous exploration of this complex parameter space so as to overcome local minima. However, this has yet to be tested for 2D galaxy images being fit with multi-components using an MCMC algorithm. In order to better quantify the systematic and random uncertainties we have performed two tests: i) tests using synthetic galaxies and ii) using real galaxies.

Synthetic galaxy generation is an ideal way of to test what can be measured in current obser-

vations as well as to assess biases in the estimated values. With a sample of computer generated galaxies one can gain information about systematic and random uncertainties in photometric decompositions using various fitting procedures, while learning more about the fitting process as well. Previous studies have also utilised this idea for both single component (Häussler et al. 2007; van der Wel et al. 2012; Newman et al. 2012) and multi-component photometric decompositions (Méndez-Abreu et al. 2008a; Davari et al. 2014; Bruce et al. 2014).

Despite the advantages synthetic galaxies have, they still lack the complexity of real galaxies. Therefore before any scientific analyses can be done, we need to verify that the algorithm can reproduce and match the parameter estimates given by other codes one the same images. Comparisons of this nature will help reveal differences between the codes as the images and systematics will be the same, and it will also highlight how different systematics will alter the fitting outcomes.

This section describes the use of a synthetic galaxy imager to test the robustness of PHI and limitations of the method. The imager simulates observations with higher complexity than has been done before. Section 3.2 outlines the detailed approach of simulating galaxy images; from the cosmological laws used describe the radial surface-brightness profiles, to the use of spectral energy distributions (SEDs) to obtain magnitudes in specific photometric bands as well as the inclusion of realistic CCD noise. Section 3.3 presents the results from fitting the synthetic galaxy sample. We then run PHI and a sample of SDSS galaxies. In Section 3.5 we show how we have used posterior predictive checks and Bayesian model comparisons to validate the final outputs. The chapter ends with a summary describing the results.

3.2 Generating synthetic galaxies

The primary use for a synthetic galaxy generator is to understand the limitations of photometric decompositions and the fitting procedure as a whole. When we apply PHI to real galaxies, the uncertainties on the parameter estimations can have a mix of causes. The errors can be associated with the systematics of the observations or they can be due to more physically motivated factors such as the size of the galaxy. By generating synthetic galaxies, with all the variables known, we can begin to understand where the algorithm fails and why.

The procedure we take for generating synthetic galaxies takes a bottom up approach, whereby a series of parameters are input to then calculate the dimensions and magnitudes of the galactic components. The code then goes on to fill pixel by pixel a corresponding intensity value according to a user defined model. Also various additions make the images more realistic such as cosmo-

logical effects associated with: different spatial resolution sampling, surface-brightness dimming, K-correction and magnitude/mass evolution due to a changing stellar population model. The various steps are described in detail in this section.

3.2.1 Cosmological background

Modern cosmological principles equip us with some essential formulae for interpreting observations and building models. If we want to understand the formation and evolution of galaxies we have to understand what is really being observed. This section (and the sections that follow) describe the essential formulae for interpreting cosmological observations. Since we will mainly be considering the post-recombination epoch, a matter dominated ΛCDM universe with $\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$ and $H_0 = 70 \mathrm{kms}^{-1}\mathrm{Mpc}^{-1}$, will be used, unless otherwise stated.

If one was to try and measure the distance to a galaxy with a tape measure, the distance you measured would continuously increase as the Universe expanded. A tape measure which extends at an infinite speed would be needed to find the proper distance to the galaxy or you could halt the expansion at its current scale factor. Beyond science fiction, neither of these possibilities are physically possible.

A procedure to measure the distance to a galaxy using observed quantities needs to be developed. Within our galaxy the trigonometric method of using parallaxes is a common procedure for calculating distances to stars. However, to measure the angular shift of an external galaxy is extreamly difficult with current instruments. Hubble's law, named after the famous Astronomer, is the name given to the observation that local galaxies have their recessional velocity approximately proportional to their distance. This idea is derived from the idea of the Hubble flow, whereby galaxies are moving away from us in relation to the expanding Universe. The law is often expressed as

$$z = \frac{H_0}{c} d, \tag{3.1}$$

where d is the distance to the object, c is the speed of light, H_0 is the Hubble constant. However, we now know that with this method, we severely underestimate the distances to galaxies at large distances due to the expansion of the Universe.

If we observe an object with a certain flux, the luminosity distance, D_L, is the distance that the object appears to have, assuming the inverse square law for the reduction of light intensity with distance holds. However, this is not the case as the Universe is expanding and for the inverse law

to hold the Universe needs to be spatially flat. The best way to visualise this scenario is to have a radiating source in the centre of a sphere with co-moving radius r_0. We observe the flux from this source on the surface of the sphere. The physical radius of the sphere is a relation involving the expansion parameter of the Universe, a_0, as well as the co-moving radius.

If space was static, then the radiation flux observed would simply be $S = L/a_0^2 r_0^2$, but the fact that the Universe is expanding affects how the photons propagate from the source to the observer. Redshift effects the flux density of the photons in three ways;

- Photon energies and arrival rates are redshifted, reducing the flux density by $(1+z)^2$,
- The bandwidth $d\lambda$ is increased by a factor of $(1+z)$, so the energy flux per unit bandwidth goes up by $(1+z)$,
- The observed photons at some wavelength λ_o were emitted at $\lambda_o/(1+z)$.

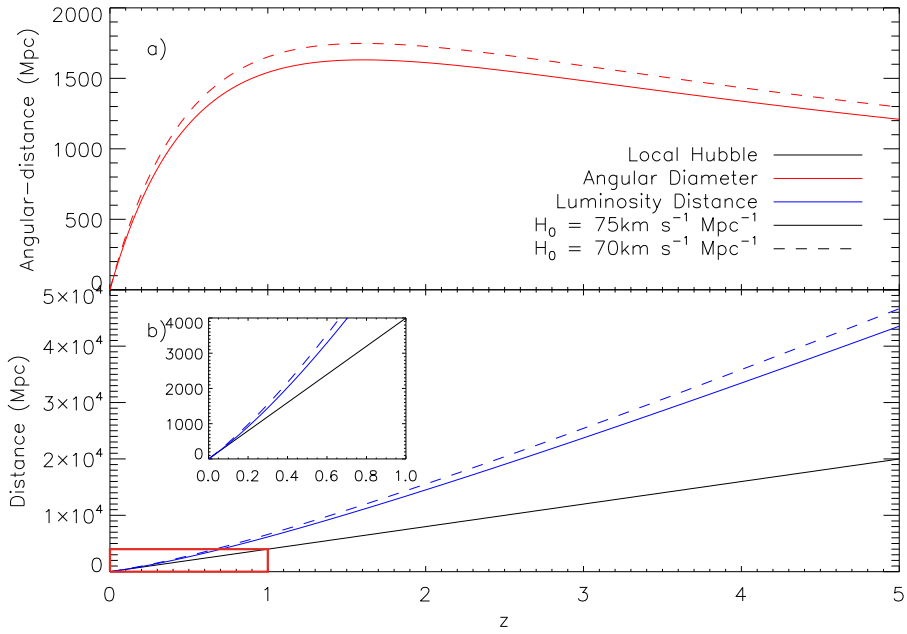


Figure 3.1: Cosmological distance calculations with the top panel showing the angular diameter distance and the luminosity distance on the bottom panel. The solid line represents a Hubble parameter of $H_0 = 75\text{kms}^{-1}\text{Mpc}^{-1}$ and the dashed line representing $H_0 = 70\text{kms}^{-1}\text{Mpc}^{-1}$. The inset shows a zoom-in at a lower redshift range.

Combining these effects and calculating the total flux measured in any bandpass we can obtain a relation for the received flux;

$$f_{\lambda_o} = \frac{L_\lambda(\lambda_o(1+z)^{-1})}{4\pi D_L^2(1+z)}, \tag{3.2}$$

where L_λ is the luminosity in units of wavelength and,

$$D_L = (1+z)a_0 r_0. \tag{3.3}$$

We can begin identify the importance of properly calculating the cosmological distance when we compare the results from using the Hubble law and the luminosity distance. Figure 3.1 illustrates this comparison, where we show that in the lower redshift range, the Hubble law and the luminosity distance are similar, but as we extend to higher redshifts ($z \gtrsim 0.1$) the two relations deviate from one another, supporting the need for the cosmological tool kit.

The luminosity distance is not the only distance measure in cosmology that can be computed using the observed properties of a galaxy. The angular diameter distance is a measure of how large an observed object appears (Raine & Thomas, 2001). An analytical definition is the ratio between the physical size, l, (in SI units) to its angular size,(θ) (in radians) (Hogg, 1999);

$$D_A \equiv \frac{l}{\delta\theta}. \tag{3.4}$$

This function is called the angular-diameter distance and has a cosmological relation given by;

$$D_A = (1+z)^{-1} a_0 r_0, \tag{3.5}$$

We can see how the angular-diameter distance changes with redshift in Figure 3.1, as well as the slight differences when we use different values for the Hubble parameter. We now have a basic understanding of how light propagates through the Universe and how we might use our knowledge of this to determine distances, but what do we actually measure and how can this be implemented into the creation of mock galaxies? The next section will describe the use of spectral energy distributions in combination with the cosmological tool kit to gain accurate values of the magnitudes of objects in the Universe.

3.2.2 Magnitudes and photometric K-corrections

In the previous sections, we have outlined the difficulties of interpreting the information we obtain from radiating sources in the Universe. One challenge is that sources observed at different redshifts are sampled at different rest-frame frequencies. The term *K-correction*, describes the transformations between the observed and rest-frame broad-band photometric measurements (Humason et al. 1956;

Oke & Sandage 1968; Hogg et al. 2002). This section follows that of Hogg et al. (2002) in describing the photometric K-correction.

If we again consider a source observed at redshift z, the observed photons have a frequency ν_o which were emitted by the source at frequency ν_e with

$$\nu_e = \nu_o(1+z). \tag{3.6}$$

The observed flux of the source is measured through a finite observed-frame band-pass with a transmission value (photon detection probability) $R(\nu)$, and the absolute (or intrinsic) luminosity is measured through the corresponding finite emitted-frame band-pass $Q(\nu)$ where $Q(\nu(1+x)) = R(\nu)$. Relating these two terms is the role of the K-correction.

The use of the K-correction takes its place in the distance modulus equation, with a source having an apparent magnitude m_R when measured in the photometric band-pass $R(\nu)$, with an absolute magnitude M_Q in the emitted-frame band-pass Q:

$$m_R = M_Q + 5\log_{10}\left[\frac{D_L}{10\text{pc}}\right] + K_{QR}, \tag{3.7}$$

where D_L is again the luminosity distance (Hogg, 1999; Hogg et al., 2002) and $1\text{pc} = 3.086\times10^{16}\text{m}$. The K-correction for this source is given by K_{QR}.

The apparent magnitude of the source is given by a relation involving its spectral density of flux f_ν, which is the energy per unit time per unit area per unit frequency:

$$m_R = -2.5\log_{10}\left[\frac{\int \frac{d\nu_o}{\nu_o} f_\nu(\nu_o)R(\nu_o)}{\int \frac{d\nu_o}{\nu_o} C(\nu_o)R(\nu_o)}\right], \tag{3.8}$$

where the integrals are over the observed frequencies ν_o; $C(\nu)$ is the spectral flux density for an object with zero magnitude. In the AB system (Oke & Gunn, 1983), is a hypothetical source with $C_\nu^{AB} = 3631\text{Jy}$ (where $1\text{Jy} = 10^{-26}\text{Wm}^{-2}\text{Hz}^{-1} = 10^{-23}\text{erg cm}^{-2}\text{s}^{-1}\text{Hz}^{-1}$) at all frequencies ν; and $R(\nu)$ describes the band-pass.

The absolute magnitude takes a similar form to the apparent magnitude but it is a measure of the brightness if the object was 10pc away, in the rest-frame (not redshifted), and compact. It is a

function of the spectral density of the luminosity L_ν (energy per unit time per unit frequency):

$$M_Q = -2.5\log_{10}\left[\frac{\int \frac{d\nu_e}{\nu_e}\frac{L_\nu(\nu_e)}{4\pi(10pc)^2}Q(\nu_e)}{\int \frac{d\nu_e}{\nu_e}C(\nu_e)Q(\nu_e)}\right], \tag{3.9}$$

where the integrals are over the rest-frame or emitted frequencies ν_e, and $Q(\nu)$ describes the band-pass Q. The functions of $Q(\nu)$ and $R(\nu)$ can describe the same filter.

Thus we can use equations 3.8 and 3.9 and substitute them into equation 3.7 to obtain a relation for the K-correction between the two band-passes:

$$K_{QR} = -2.5\log_{10}\left[(1+z)\frac{\int \frac{d\nu_o}{\nu_o}L_\nu(\nu_o(1+z))R(\nu_o)\int \frac{d\nu_e}{\nu_e}C(\nu_e)Q(\nu_e)}{\int \frac{d\nu_o}{\nu_o}C(\nu_o)R(\nu_o)\int \frac{d\nu_e}{\nu_e}L_\nu(\nu_e)Q(\nu_e)}\right]. \tag{3.10}$$

The above calculations are performed in frequency units, but we can convert to units of wavelength quite simply by using the following relations:

$$d\nu f_\nu(\nu) = d\lambda f_\lambda(\lambda), \tag{3.11}$$

$$\lambda\nu = c. \tag{3.12}$$

Thus to compute the K-correction, the source flux density needs to be accurately known, as well as the standard-source or zero-point magnitudes C and the bandpass functions. We are now well equipped with a set of tools that will enable us to perform some magnitude calculations, this will become important in the next stages of the mock galaxy build.

3.2.3 Stellar populations and magnitude determination

In order to create our synthetic galaxy images as realistic as possible, we simulate the process of computing the magnitudes from observations, *i.e.,* by observing the flux densities in specific band-passes. To do this we include a set of stellar population models with simulated spectral energy distributions (SED) and known star formation histories (SFH) into the generation process.

Here we use the GALAXEV library of evolutionary stellar population synthesis models that are computed using the isochrone synthesis code of Bruzual & Charlot (2003). Where a wide range of

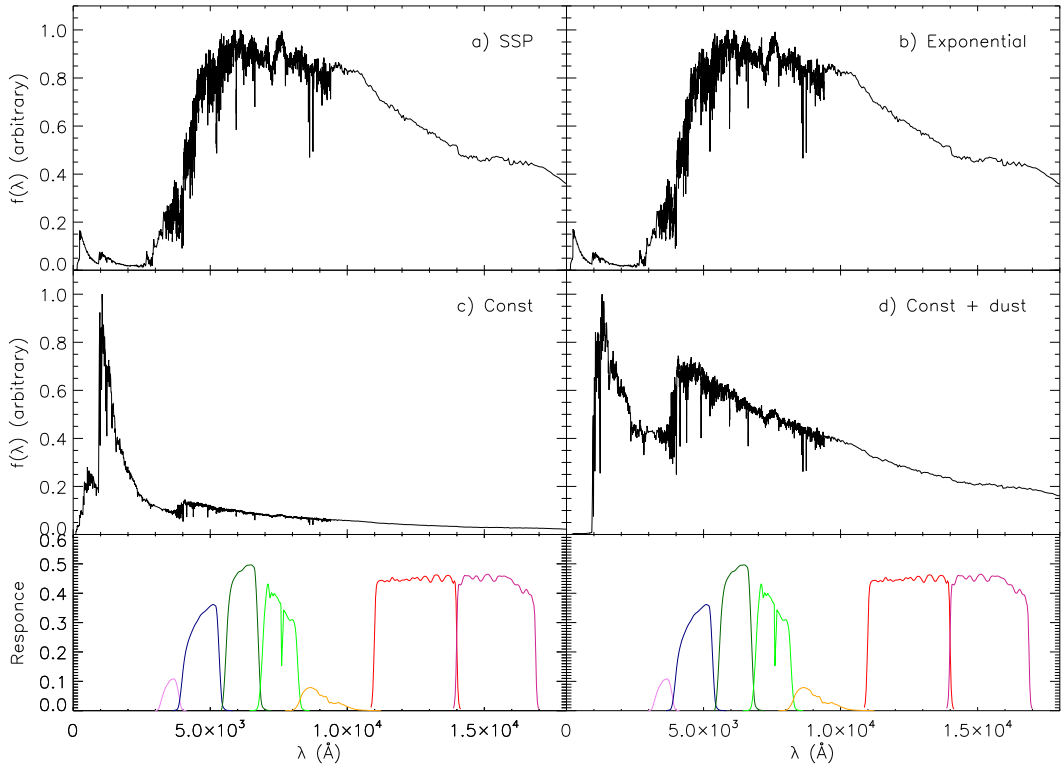


Figure 3.2: A sample of the stellar population synthesis models used to determine the apparent magnitude in the band-passes. The four starformation histories used are a) a steep stellar population, b) an exponentially declining SFH, and c) constant starformation, and finally d) constant with dust attenuation. Each SED has a formation redshift of $z_f = 10$, with the flux density shown for the most local of the sample. The trasmission curves for the u, g, r, i, z, J, and H band-passes for SDSS and CANDELS are shown in the bottom panel.

ages, metalicities at a resolution of 3Å across the wavelength range 3200Å to 9500Å, with lower resolution outside this range were used. Four types of SEDs have been chosen for the use in the imager. The models have different SFHs: (1) an exponentially declining star formation rate (with inverse decay rates between 0 and 1 Gyr^{-1}), (2) a single burst with star formation rate thereafter, and (3) a constant star formation rates (SFR).Dust attenuation is optionally included in the constant SFH model at a level of $\tau_V(\text{young stars}) = 1$ and $\mu = 0.3$, *i.e.,* the optical depth of stars younger than 10^7 yrs is 1, and stars older than 10^7 yrs are attenuated only by diffuse interstellar dust with a τ_V of 0.3 (for more information on the attenuation law see Wild et al. 2007). Figure 3.2 shows a sample of the SEDs used in the modelling of the galaxy images, with the single stellar population (SSP), the exponential, and the constant and constant with dust models. Also shown are the band-passes, these are the set predominately used in the modelling, but the code can simulate whatever set of band-passes from a range of telescopes, those presented here are from the Sloan Digital Sky Survey (u, g, r, i, and z bands) and two bands from the CANDELS survey on HST (the F125W and

the F160W bands).

We can now include the equations presented in the previous section to calculate the magnitudes in the band-passes. The synthetic galaxy imager will accept either an input for the absolute magnitude of a galaxy or a list of galaxy absolute magnitudes. The user will first choose a specific SFH as well as a redshift at which the stars were formed, the stellar formation epoch. This does not correspond to the formation of the galaxy, as the model assumes the objects are isolated without any merging happening. The code will then choose a sample of SEDs, that once their age is added to the formation epoch, are yopunger than the age of the Universe. The redshift of the model galaxy is then converted to a look-back time, and the SED identified that corresponds to that epoch.

As the SEDs are simulated for a $M_* = 1M_\odot$ galaxy at $z = 0$, they need to be scaled to match the input absolute magnitude. With the knowledge of the transmission of the filters, equation 3.9 gives us the absolute magnitude relevant to the SED, but not to the simulated galaxy wanted by the user. The magnitude then needs to be scaled to the input absolute magnitude using the relation;

$$M_{in} - M_{calc} = -2.5 \log_{10}(x), \tag{3.13}$$

rearranging we obtain

$$x = 10^{0.4(M_{calc} - M_{in})}, \tag{3.14}$$

where x is the flux ratio between two related quantities to M_{calc} and M_{in}. Using this scaling factor, (which can be thought of as giving the galaxy a specific mass) we can scale the SED to obtain the correct flux density in the bandpass that corresponds to the input absolute magnitude. Alternatively, the user can input a desired stellar mass for the synthetic galaxy to be used as the scale factor. Once the SED has been scaled, the apparent magnitude in either the same bandpass or any other bandpass can be calculated using equation 3.8.

3.2.4 Surface-brightness distributions

The surface-brightness distributions (SBD) are used to describe how the surface brightness varies over the component modelled. The Sérsic profile, is commonly used to describe elliptical and bulge light distributions and the exponential profile is used to describe that of the disc.

The Sérsic power law is one of the most frequently used functions to study galaxy morphology, and has the following form, providing the intensity at some radius R:

$$I(R) = I_e \exp\left[-b_n\left[\left(\frac{R}{R_e}\right)^{1/n} - 1\right]\right], \tag{3.15}$$

where I_e is the intensity at the effective radius R_e that encloses half of the total light from the model (Graham & Driver, 2005). The constant b_n is directly related to the Sérsic index n, which describes the shape of the light-profile (see Fig. 3.3). Common values for b_n are $b_4 = 7.669$ and $b_1 = 1.678$. Analytical expressions for the b_n parameter have also been issued. Capaccioli & Caon (1989) provide a analytical expression, which takes the form $b_n = 1.9992n - 0.3271$, for $0.5 < n < 10$.

One can integrate equation 3.15 over an area $A = \pi R^2$ to obtain the total luminosity, L (Ciotti, 1991):

$$L = 2\pi n I_e R_e^2 \frac{e^{b_n}}{(b_n)^{2n}}\Gamma(2n). \tag{3.16}$$

where the $\Gamma(2n)$ is the gamma function defined by

$$\Gamma(2n) = \int_0^\infty e^{-x} t^{n-1} dx. \tag{3.17}$$

and $x = b_n(R/R_e)^{1/n}$ has been substituted.

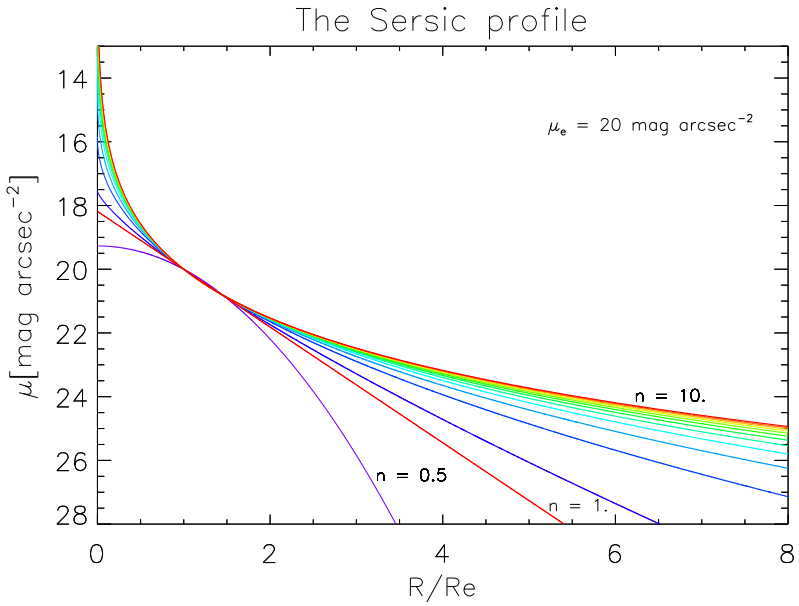


Figure 3.3: The Sérsic profile. The different coloured lines represent varying values of the Sérsic index n, where $n = 0.5$ is a Gaussian and $n = 1$ describes and exponential profile. These profiles have a effective surface brightness of $\mu_e = 20\,\mathrm{mag\,arcsec^{-2}}$.

For an exponential profile we have $n = 1$, of which 99.1% of the flux resides within the inner

$4R_e$. For the de Vaucouleurs profile ($n = 4$) 84.7% of the flux resides in the inner radius less than $4R_e$. By multiplying the negative logarithm of equation 3.15 by 2.5 gives the surface brightness profile. Figure 3.3 illustrates the SBD of the Sérsic profile for a range of n values.

Another transformation arises from the use of scale-lengths h rather than effective radii R_e. When the Sérsic model is written as

$$I(R) = I_0 \exp\Big[-(R/h)\Big], \tag{3.18}$$

where

$$I_0 = I_e e_n^b, \tag{3.19}$$

and

$$h = R_e / b_n^n. \tag{3.20}$$

This is the exponential law used to describe the SBD of the disc component (Freeman 1970; Ellis & Perry 1979; Davies et al. 1988; Graham & Driver 2005), where I_0 and h ($R_e = 1.678h$, when $n = 1$) is the central intensity and the scale length of the disc, respectively. Similarly to equation 3.16, we can integrate over the projected area of the exponential intensity profile to obtain the total luminosity:

$$L = 2\pi \int_0^\infty I(R) R \, dR = 2\pi I_0 h^2. \tag{3.21}$$

Studies of galaxy structure commonly refer to the light fraction contributed by the bulge and the disc, alternatively it is expressed as the bulge-to-total luminosity ratio (i.e., $B/T = B/(B + D)$, where B is the total luminosity from the bulge, and D is the disc luminosity). We can use equations 3.16 and 3.21 to find the bulge-to-disc ratio:

$$B/D = \left(\frac{I_e}{I_0}\right)\left(\frac{R_e}{h}\right)^2 \frac{n \exp(b_n)}{b_n^{2n}} \Gamma(2n). \tag{3.22}$$

3.2.5 Surface brightness dimming

Geometrical effects due to the galaxies being at high redshift modify the surface brightness. In a Euclidean Universe the situation would be simpler as the flux we receive from an object would follow the inverse square law, also the projected area of the object will decrease following the same law leading to a constant SBD with distance. However, in an expanding universe, we have discussed the effects that happen to the photons as they propagate through space-time, affecting the surface brightness we observe. We can calculate by what amount the brightness will decrease by if we again consider an emitting object with an extended size $\delta\theta$, linear size l, and luminosity L. The solid angle subtended by the source is $\delta\Omega = \pi\delta\theta^2/4$. We can use the relationship between linear size and angular size from equation 3.5 to show that;

$$\delta\Omega = \frac{\pi l^2}{4D_A^2} = \frac{\pi l^2(1+z)^2}{4a_0^2 r_0^2}. \tag{3.23}$$

The surface brightness of a source is thus

$$\mu_{\text{observed}} = \frac{f_o}{\delta\Omega} = \frac{L}{\pi^2 l^2(1+z)^4} = \frac{\mu_{\text{emitted}}}{(1+z)^4} \tag{3.24}$$

where μ_{emitted} is the emitted surface brightness. To implement the effects of surface brightness dimming, the imager needs to modify the physical size according to D_A (see equation 3.5).

3.2.6 Telescope and instrument systematics

We now model the seeing effects which scatter light and produce a loss of spatial resolution in real telescope images. These effects impact the central regions where the slope of the radial surface brightness profile is steeper than in the rest of the galaxy.

Point spread function: We can input three kinds of point spread functions (PSFs): A two-dimensional Gaussian PSF, a two-dimensional Moffat PSF (Figure 3.4) or a user-supplied PSF (see Figure 3.5 for an example of externally generated PSFs). The code normalises the total flux in all input PSFs to ensure flux conservation when convolved with the surface brightness from the galaxy image. The Gaussian and the Moffat PSFs are generated in the imager with the full width half maximum (FWHM) specified by the user.

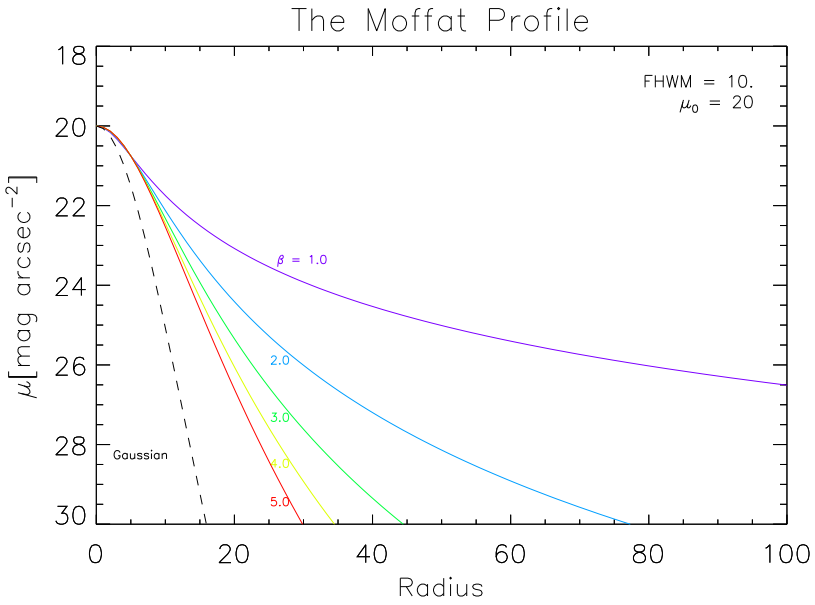


Figure 3.4: Comparison between a Gaussian PDF (black dashed line) and a Moffat PDF with differing β values. These profiles have a effective surface brightness of $\mu_e = 20\,\mathrm{mag\,arcsec^{-2}}$ and a FWHM= 10. pixels.

The Gaussian PSF is given by:

$$PSF_{gauss}(r) = \frac{1}{2\sigma^2\pi} e^{-\frac{r^2}{2\sigma^2}}, \tag{3.25}$$

where σ is the standard deviation, and related to the FWHM using FWHM $= 2\sqrt{2\ln(2)}\sigma$. The Moffat function has a similar form:

$$PSF_{moffat}(r) = \frac{\beta-1}{\pi\alpha^2}\left[1+\left(\frac{r}{\alpha}\right)^2\right]^{-\beta} \tag{3.26}$$

where the parameters α and β define the profile shape and FWHM $= 2\alpha\sqrt{2^{1/\beta}-1}$. Figure 3.4 shows how the Moffat and the Gaussian differ in the outer wings, and with increasing values of β the shape will tend to a Gaussian. The PSF is convolved with the galaxy image using a Fast Fourier Transform algorithm which reduces the computation time.

Background noise: The main sources of noise in CCD data are readout noise and photon counting. As well as this, there are cosmic ray hits, affecting a small number of CCD pixels. The readout noise is characterised by σ_B (rms in the data numbers), which is measured from bias images. We obtain sky estimates for each image from the band we wish to simulate observations in, and reject stars and cosmic rays with a sigma clip. We can then calculate the σ_B value for the sky background, which is used in the imager code to generate fake background noise modelled on real images.

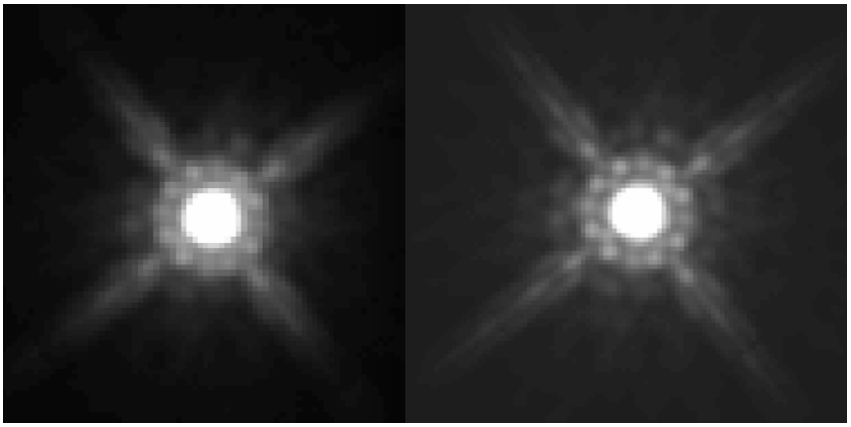


Figure 3.5: Examples of PSF images that can be entered into the imager. These are examples of simulated HST PSFs for the J (left) and H (right) band-passes from Krist et al. 2010.

3.2.7 Summary: Building a synthetic galaxy

In this section we have outlined the various relations and parameters needed to build an image of a galaxy a given SFH, star formation epoch, SBD, and where the object is placed at a redshift which is observed in a specific filter using a modern CCD. The imager provides realistic models crucial to in testing the PHI algorithm as well as to define the limits of 2D photometric decompositions from observations with specific telescopes and as a function of redshift.

3.3 Performance tests on synthetic images

In this section we use synthetic galaxy images to test the accuracy and robustness of the PHI algorithm. Synthetic galaxies lack the complexity present in real galaxies but allow us to check for statistical and systematic errors associated with the method thanks to the knowledge of the true parameter values.

In order to generate realistic galaxies, one might use known scaling relations to resemble structures in the Universe or we can simply use the results from previous galaxy surveys with complete samples. Using just a sample of synthetic galaxies created from a grid of parameter values would be unrealistic as a significant fraction would lie outside the physically realistic parameter space.

Based on the results of Gadotti (2009), in which 260 elliptical and 380 bulge + discs were analysed with single Sérsic and Sérsic + exponential profiles, respectively, we create our sample of synthetic galaxies. The parameter distributions resemble a realistic look at how local galaxy structures are distributed. We simulated the noise in the images to match that of SDSS by removing all objects from the data frames in SDSS and estimating the background noise with a Poisson noise model. We use a Gaussian for the PSF where the FWHM is again taken from the values given by

Table 3.1: Parameter ranges for synthetic single component Sérsic galaxies and two-component Sérsic + Exponential galaxies. Geometric parameter ranges are also shown.

Synthetic Elliptical	
Parameter	**[min, max]**
Sérsic magnitude ($m_{S,i}$)	[14, 17]
Effective radius (R_e in arcsecs)	[1.5, 6]
Sérsic index (n)	[1.9, 7.5]
Synthetic Bulge + Disc	
Parameter	**[min, max]**
Sérsic magnitude ($m_{S,i}$)	[15, 21]
Effective radius (R_e in arcsecs)	[0.4, 2.24]
Sérsic index (n)	[0.5, 7]
Exponential magnitude ($m_{E,i}$)	[15, 18]
Scale height (h in arcsecs)	[1.3, 7]
Geometric parameters	
Parameters	**[min, max]**
Axial ratio (q)	[0.6, 1]
Position angle (PA in degrees)	[-360, 360]

Gadotti (2009) for their SDSS images. The instrumentation parameters are motivated by SDSS DR7 images with the pixel scale (0.396 arcsec / pixel), and the typical values of the CCD gain ($4.86e-$/ADU) and read-out noise ($5.76e-$) are taken to mimic the instrumental set-up.

The galaxy images were simulated to be within the i-band of SDSS for the sake of simplicity. All the synthetic galaxy possible population parameter ranges are specified in table 5.3. The sample parameters permit a large range in the luminosity ratio between the bulge and the total luminosity of the galaxies being $B/T = [0.01, 0.8]$. With the generated synthetic galaxies we have run `PHI` as if they were real.

Figures 3.6 and 3.7 show the combined errors of the marginalised posterior distributions of all the synthetic galaxies stacked. The error for a given parameter is calculated by taking the residual and dividing it by the real values to obtain a fractional error, error $= (x_{out} - x_{in})/x_{in}$, where x_{in} is the true parameter value and x_{out} are the values given in the MCMC output. In figures 3.6 and 3.7, the marginalised error posterior are shown on the diagonals, and the joint marginal distributions are shown on the off-diagonals. The dotted black line showing the median of the distributions and the dashed line signifying the first and third percentiles.

Both the individual galaxy posterior distribution shown in Figure 2.5 and the total combined

Table 3.2: Table showing the medians of the posterior errors shown in Figures 3.6 and 3.7 along with the 16th and 84th percentiles for the synthetic elliptical and bulge + disc galaxies.

Synthetic Elliptical			
Parameter errors	**Median**	**16%**	**84%**
$\Delta I_e/I_{e,0}$	0.003	-0.022	0.049
$\Delta R_e/R_{e,0}$	0.004	-0.021	0.018
$\Delta n/n_0$	-0.004	-0.033	0.012
Synthetic Bulge + Disc			
Parameter errors	**Median**	**16%**	**84%**
$\Delta I_e/I_e$	0.005	-0.25	0.22
$\Delta R_e/R_{e,0}$	-0.001	-0.15	0.14
$\Delta n/n_0$	-0.014	-0.19	0.05
$\Delta I_0/I_{0,0}$	-0.002	-0.09	0.04
$\Delta h/h_0$	0.0001	-0.02	0.03

posterior errors presented in Figures 3.6 and 3.7 show that there are clear covariances between the model parameters and their errors. The origin of this covariance can be thought of as the following. For a given surface brightness profile, if the effective intensity is overestimated (underestimated) the effective radius will be overestimated (underestimated) to compensate for the observed surface brightness. Furthermore, as the effective radius is overestimated (underestimated) the Sérsic index will also be overestimated (underestimated) due to the compensation needed to increase the concentration of intensity within a now larger effective radius.

Table 3.2 present the statistics of Figures 3.6 and 3.7. It can be seen that the level of systematics for both the synthetic ellipticals and bulge + disc galaxies is minimal and the 1σ errors (16% and 84% percentiles) are usually below a 20% error (less than 1% for most of the parameters). Uncertainties also agree with the literature error estimates for similar artificial images (see (Bruce et al., 2012)). This suggests that the code is robust and reliable.

Using the simulated galaxies we investigate the effect of observational attributes such as the Sérsic index n on the various profile parameters as well as their effect on n itself. We also identify the errors on another common parameters used for galaxy classifications, the bulge-to-total ratio, B/T. The end of this section includes information on the error estimates of the disc (or outer component) parameters.

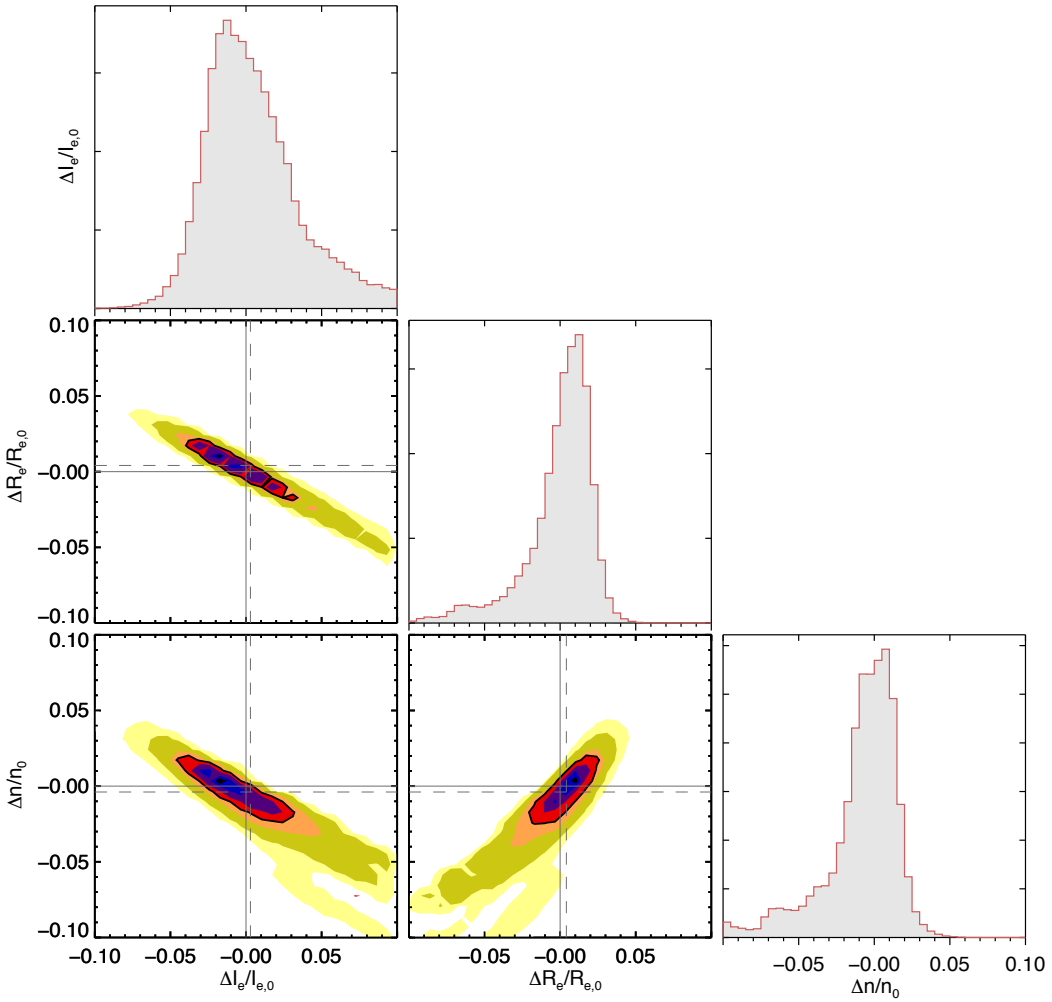


Figure 3.6: Posterior error distribution for the entire ensemble of synthetic elliptical galaxies. The full marginal error distribution for each structural model parameter is shown on the diagonal. Joint marginal pairs of parameters are shown on the off-diagonal. The seven color contours represent the 10, 30, 50, 68, 80, 95, and 99% confidence levels and the black solid line signifies the 68% confidence level. The solid grey line shows the true value of this synthetic galaxy and the dashed line represents the median of the stacked posterior distribution. All the parameter values have some degree of correlation with the most correlated parameters being the Sérsic profile parameters between each other.

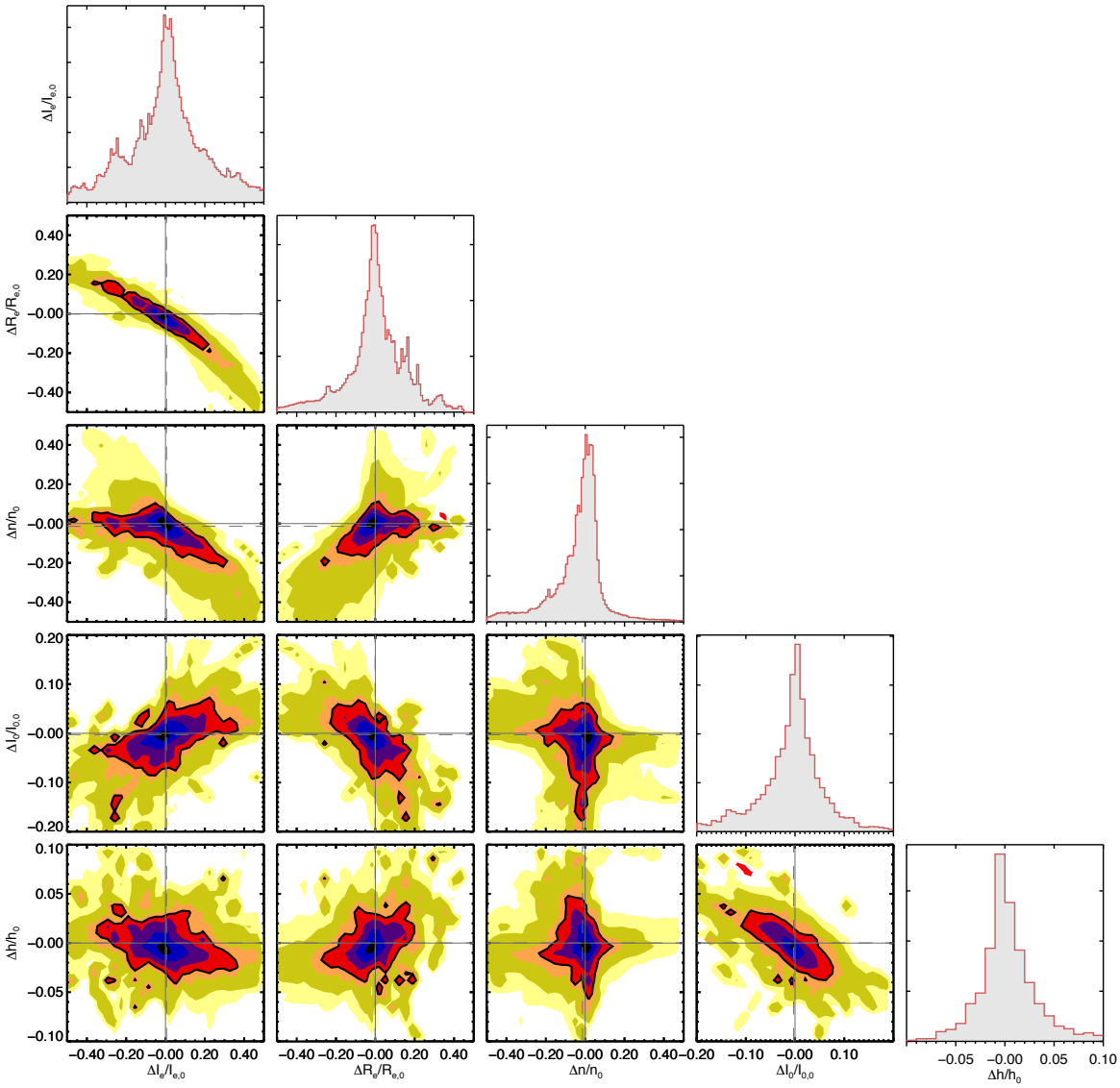


Figure 3.7: Posterior error distribution for the entire ensemble of synthetic bulge+disc galaxies. Lines and contours are the same as figure 3.6.

3.3.1 Impacts of and on the Sérsic index, n

To investigate the trends with the Sérsic index n, we used the results presented in Figure 3.7 to identify the major contributors to the uncertainty in n as well as how the true value of n affects other parameters in the model. This analysis is performed on two-component synthetic galaxies. In Figure 3.8 we binned a selection of input parameters to show what implications they have on the estimation of n via box plots. The box encloses the interquartile range (IQR), with the median shown with the horizontal line within the box. The whiskers extend out to the maximum or minimum values of the data. Figures 3.8 to 3.10 show the fractional errors for various model parameters vs. the true parameters of the synthetic galaxies. The marginalised error posterior for the Sérsic index has strong correlations with the rest of the parameters present in the Sérsic surface brightness profile (also seen in Fig. 3.6 and 3.7). Paying close attention to the median line, we can see that the deviation from the true answer is on average less than 5%, thus showing that the code can recover the parameter values robustly without a significant bias. In 3.8 we can identify that not only do the individual parameters of the Sérsic profile have an effect on n but there is also a systematic effect caused by the decreasing R_e to PSF ratio. Once the PSF begins to dominate over the effective radius the IQR increases, although the median line remains very close to 0. The IQRs are sensible throughout with the error on n only ever reaching $\Delta n = +/-0.6$.

We can also identify any biasing effects the Sérsic index may have on the other model parameters. Figure 3.9 shows the fractional error of the structural parameters: I_e, R_e, I_0, and h vs. the input values of n. We find that there is evidence of a biasing effect on the error of I_e and R_e when $n > 5$. The origins of this bias could be due to the low number of synthetic galaxies in the later bins, hence, a low number of bad fits could result in an overall bias in that bin. After searching for any repetitions of such biases it appears that this only happens for the Sérsic surface brightness structural parameters vs. the input n. This is also suggested by the covariances seen in Figures 3.6 and 3.7. We found that for a higher Sérsic index the algorithm takes longer to converge, also implying difficulties in this region of the parameter space. Further synthetic galaxies in the larger Sérsic index bins could be used to determine whether there is actually an inherent bias within the algorithm, or the surface brightness profile models.

3.3.2 Effects on B/T

A common representation of galaxy morphology is given by the B/T ratio. We investigate any trends in the B/T due to all the model parameters being present in the calculation of the B/T. One might

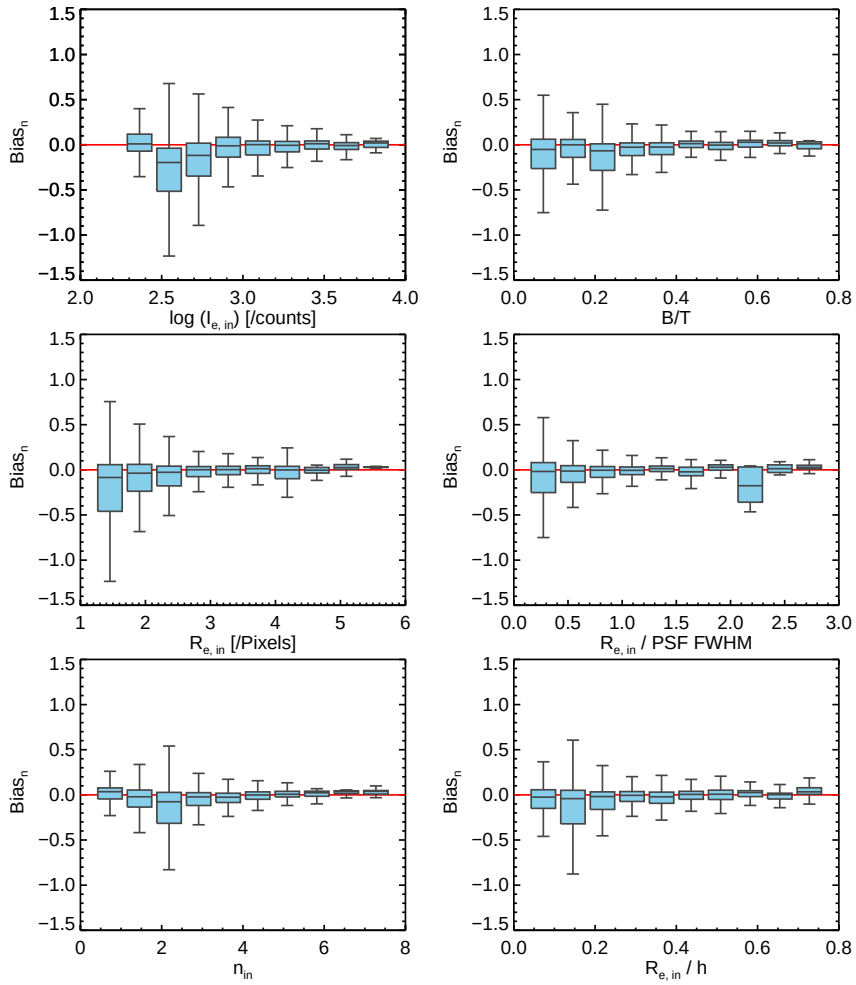


Figure 3.8: Box-plots representing the posterior errors of the n vs. binned true input values for I_e, R_e, n, B/T, the ratio between R_e and the PDF FWHM, and the ratio between R_e and h. The box limits represent the 16 and the 84% percentiles and the median values for each bin are shown in the horizontal line cutting each box. The whiskers show the extent of the distributions in each bin.

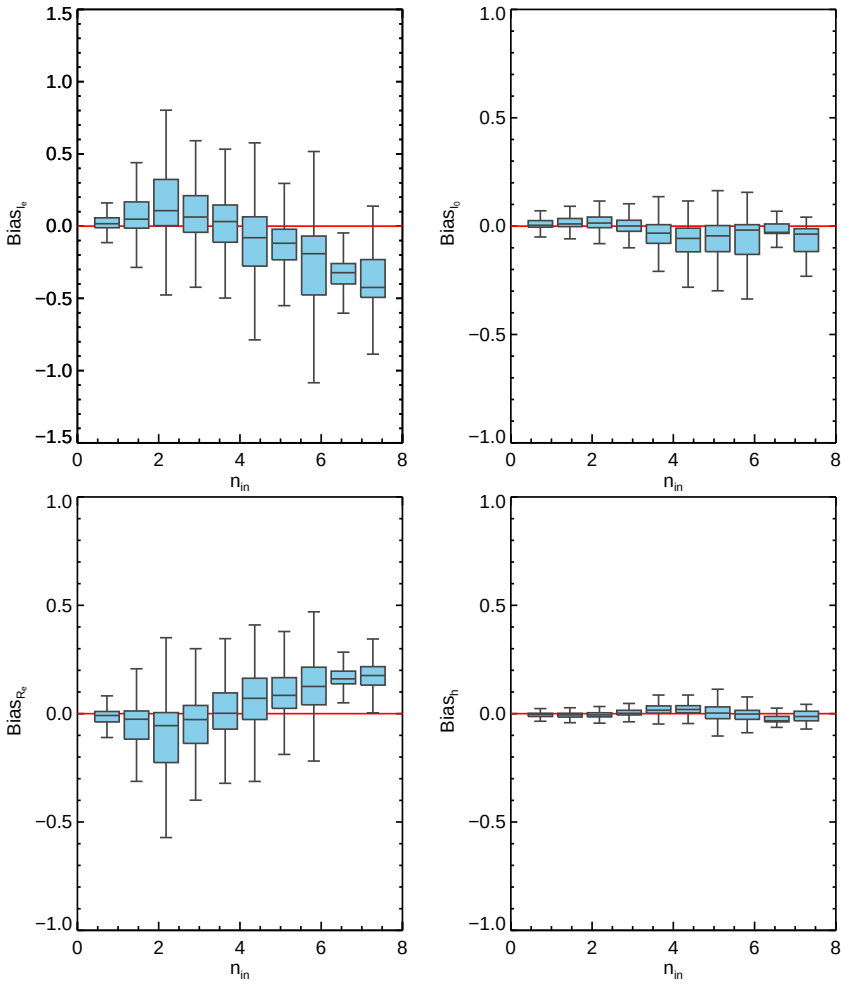


Figure 3.9: Same as Figure 3.8 but showing the posterior errors of I_e, R_e, I_0 and h vs. binned true input values for n.

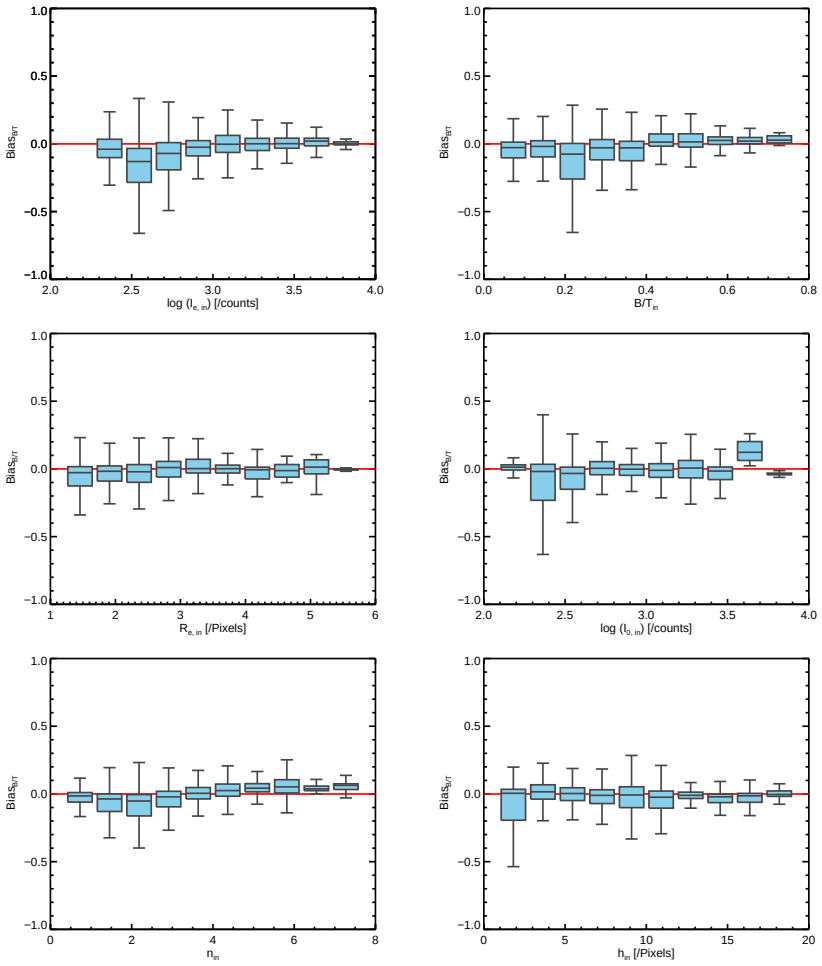


Figure 3.10: Same as Figure 3.8 but showing the posterior errors of the B/T vs. binned true input values for I_e, R_e, n, B/T, I_0 and h.

suggest that if any systematic biases or significant errors in the fitting of the model they will be present in the marginalised posteriors of the B/T.

The top right panel of Figure 3.8 shows that there is no clear evidence that the input B/T has any effect on the Sérsic index parameter as well is its uncertainties. Figure 3.9 shows that for the effective intensity and the effective radius are biased at higher values of n. This slight bias in the I_e and R_e may have a subsequent affect on B/T. Figure 3.10 suggests that the true parameters do not affect the final B/T distributions significantly for any value of input parameter. The final estimates of B/T appear stable against other parameters and errors.

3.3.3 Effects on the Exponential parameters

Figure 3.11 shows the fractional errors for the disc structural parameters, with the right column in Figure 3.9 showing how the Sérsic index effects for the same parameters. As the exponential profile is commonly related to the disc of two-component galaxies we will refer to this as the disc for simplicities sake.

Overall there appears to be little to no bias on the disc parameters with the exception of the input n. The right panels of Fig. 3.9 show that when the Sérsic index increases the median values for the errors of I_0 and h become negatively and positively biased, respectively. Although, the fractional errors on h are minimal due to the disc being a well resolved component. The origins of this may also explain the growing error regions when the inner component becomes larger in relation to the outer (bottom right panel in Fig. 3.8). When a galaxy has a dominant disc the inner component becomes increasingly saturated as the light from the disc blurs more and more into the light from the bulge.

3.4 Applications to real galaxies

The next step is to apply PHI to a sample of real galaxies in the local Universe to assess the functionality and robustness of the method in a fully realistic scenario. We perform a two tiered approach where (1) we compare the results from fitting the same galaxy images between two different codes (with two different minimisation schemes) and (2) we compare different images of the same galaxies with two different codes again. The aim of tier (1) is to identify any differences between the codes as the images and systematics will be the same. Tier (2) will mainly highlight how different systematics will alter the fitting outcomes.

We utilise public Sloan Digital Sky Survey (SDSS; Strauss et al. 2002) images that have been previously analysed in Gadotti (2009). This analysis was done on ~ 1000 nearby galaxies using the BUDDA code (de Souza et al., 2004) to perform bulge/disc/bar 2D photometric decompositions. The stellar masses are above $10^{10} M_{\odot}$ and within a redshift range of $0.04 \leq z \leq 0.06$. We have also removed those galaxies that host a bar. For this study we want to focus purely on the one and two-component galaxies namely the galaxies in Gadotti (2009) classified as elliptical galaxies and galaxies with a bulge and a disc. The sample used in this paper contains 250 elliptical and 350 bulge+disc galaxies where the concentration parameter (C) was used to separate the two populations. (see Gadotti, 2009). Elliptical galaxies were defined with $C > 3$ with disc-dominated galaxies

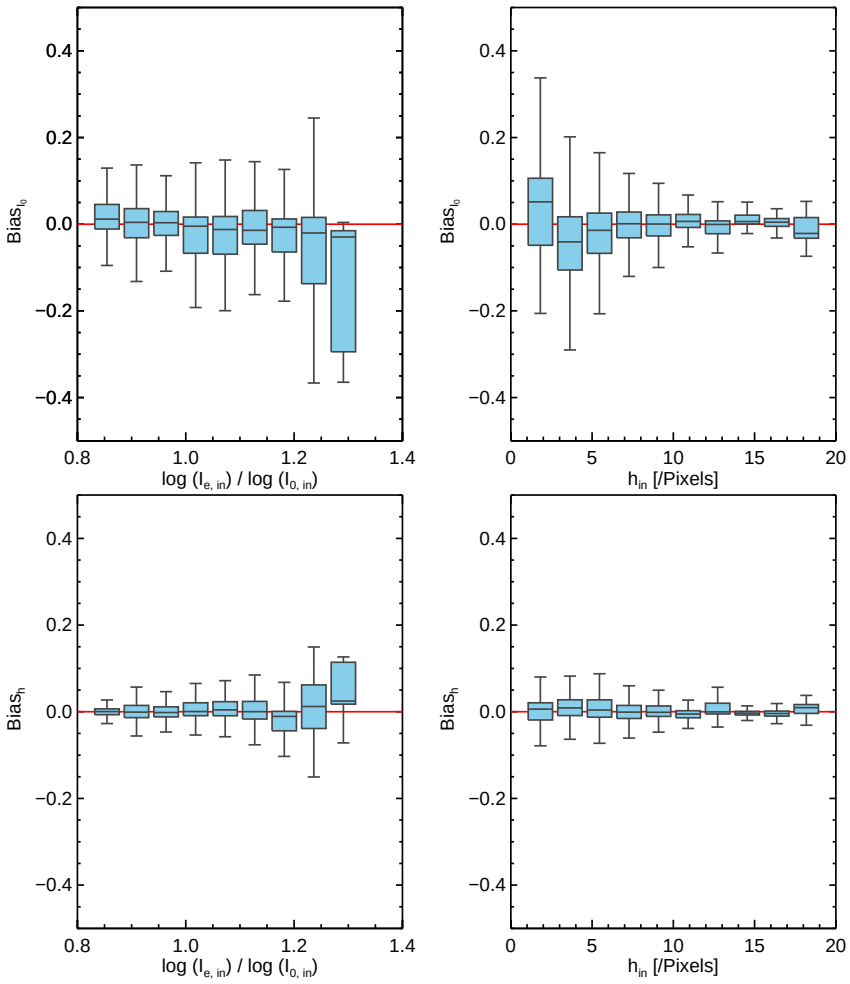


Figure 3.11: Same as Figure 3.8 but showing the posterior errors of I_0 and h vs. binned true input values for B/T, I_0 and h.

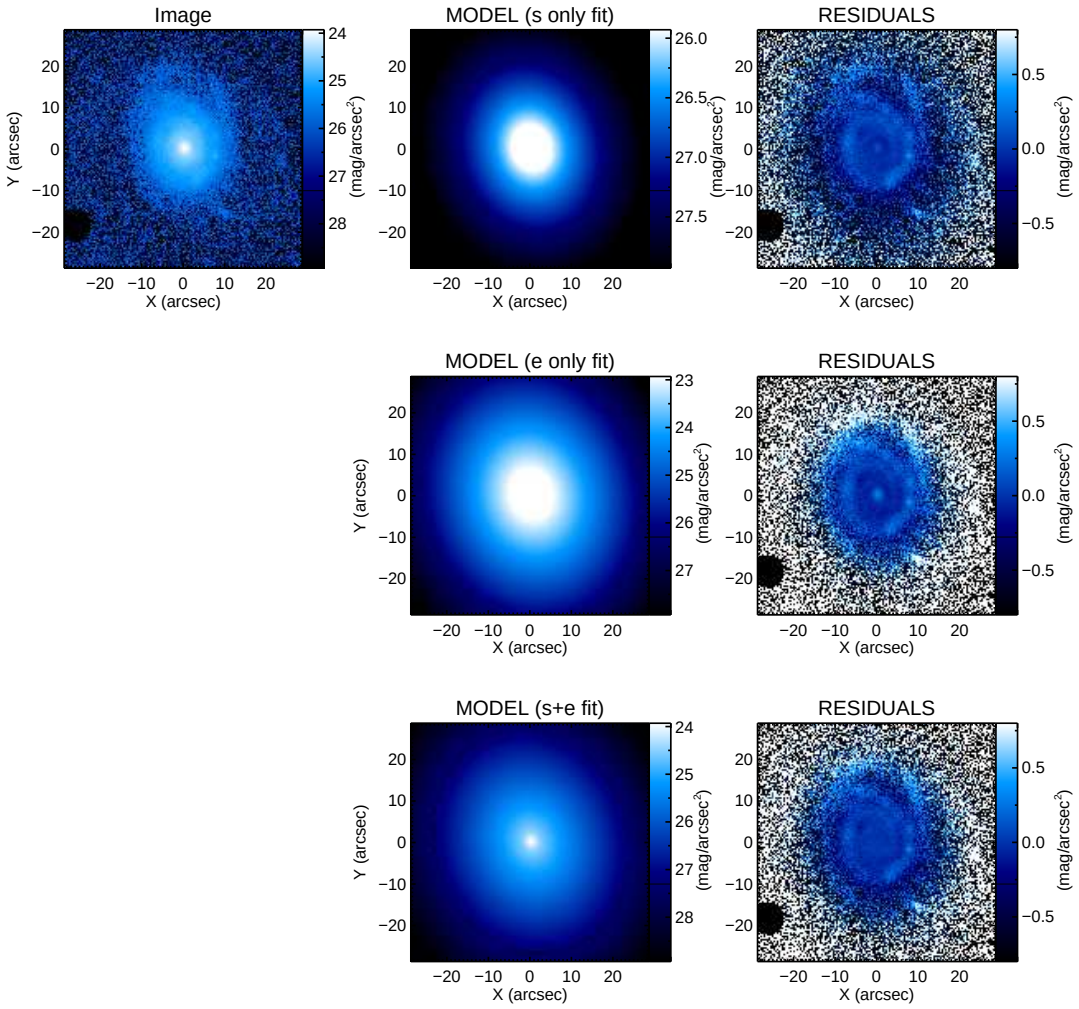


Figure 3.12: Results from a MCMC decomposition for a galaxy from the SDSS sample. The top row shows the data (left), the Sérsic only model fit (middle) and the residual. The second row shows the exponential only fit and its corresponding residual. The bottom row shows the bulge-disc model fit and its corresponding residual. All the models were made using the medians from the posterior distributions.

having $C < 2.5$. The imaging used to classify the sample and perform the 2D photometric decompositions in Gadotti (2009) was based on the SDSS data release 2 (DR2; Abazajian et al. 2004). The fits were performed on sky-subtracted SDSS images. The sky level could be fit within PHI but as previously mentioned in Section 2.2, Méndez-Abreu et al. (2008a) provided evidence on how small errors in the subtraction of the sky background can have significant effects on the fitting of the image thus supporting the need to do the process separately. By masking out the galaxy and calculating the mean and standard deviation of the sky we have verified that the sky had been successfully subtracted from all the images.

For this study we use the DR7 (Abazajian et al. 2009) images with PSFs obtained by fitting nearby stars to each object with a Moffat function to derive the full width half maximum (FWHM) and β parameter. Variability in the star sample is limited to the 50% percentile of a sample size between 5 to 10 stars in the immediate vicinity of the galaxies. Segmentation maps were created using a similar approach to that Source Extractor (Bertin & Arnouts 1996) where a group of pixels are identified as belonging to the source of interest and separated from any other contaminating nearby objects. As the objects were originally selected based on their isolation, overlapping flux sources are not a problem. As described in Gadotti (2009), despite the constraints imposed in the sample selection for the axial ratio and concentration parameter, we consider the final sample to be a fair representation of the galaxy population in the local Universe. We have run the PHI algorithm fitting each galaxy with both a single Sérsic and a Sérsic+exponential model to be able to do a proper model comparison (see section 3.5). The output posterior probabilities have been created with three simultaneously running chains and had the burn-ins removed.

In order to perform the double test proposed previously (tier 1 and 2), the sample of ellipticals and bulge+disc galaxies have been analysed using GASP2D (see Méndez-Abreu et al. 2017 for an application to large samples). It is important to note that GASP2D and BUDDA share the same minimisation engine, a Levenberg-Marquardt method, but the image analysis using GASP2D has been performed using the same images, PSFs, weight maps, and masks as used in PHI.

3.4.1 Comparison of Elliptical galaxies

Figure 3.13 compares the estimated Sérsic profile parameters of the elliptical galaxies classified in Gadotti (2009). The blue circles show the difference between PHI and GASP2D and the red diamonds is the difference between the results of PHI and those of Gadotti (2009). We show only those galaxies where all chains have successfully converged in the sampling phase as well as removing any catastrophic failures that occurred when running GASP2D (a total of 20).

We focus first on the comparison between PHI and GASP2D where the same images are fit with the same PSFs, weight maps, and segmentation maps. There is a subtle deviation from an exact one-to-one match for higher R_e and n. In GASP2D, a run is determined to have reached the global minimum and give its final answer when the deviations between the χ^2 of two consecutive iterations is lower than a given threshold. For larger n, R_e combinations, the variations in χ^2 are insignificant compared to when n and R_e are lower. Due to the exponential nature of the Sérsic profile changes in n at the lower end (*i.e* ≤ 2) have a higher impact on the surface brightness whereas changes made

at the higher n have a lower impact and thus the error region would be larger. Otherwise the codes appear to have a tight correlation and agree on average with standard deviations for the parameter residuals are; $\sigma_{log(I_e)} = 0.102$ counts, $\sigma_{R_e} = 2.01$ pixels and $\sigma_n = 0.488$.

When we compare the output posterior medians to the results obtained by Gadotti (2009) we see some significant differences between all the parameters. Where Gadotti (2009) had tighter distributions in all the parameters the results we obtain with PHI span larger ranges. For the distribution of n in Gadotti (2009) has a mean of 3.77 and standard deviation of 0.87, compared to our mean of 4.05 and standard deviation 1.17. The difference between the two analyses also has a strong function of the structural parameters, with PHI finding smaller I_e in low intensity objects and larger sizes and Sérsic indexes for large, high-n objects.

3.4.2 Comparison of bulge+disc galaxies

Figure 3.14 compares the estimated parameter values for the sample of bulge+disc galaxies. The comparison between the posterior medians and the results of GASP2D for the Sérsic profile parameters agree, although with larger scatter than for the one component galaxies. The standard deviations for the parameter residuals are; $\sigma_{log(I_e)} = 0.36$ counts, $\sigma_{R_e} = 2.00$ pixels, $\sigma_n = 1.89$ $\sigma_{log(I_0)} = 0.09$ counts, $\sigma_h = 1.46$ pixels.

The comparison between the posterior medians and the results of Gadotti (2009) shows a similar trend to that of Figure 3.13 where the Sérsic parameters are consistently offset from one another, particularly at large R_e and small I_e. There is a tighter correlation for the disc parameters however suggesting the fitting of an exponential component to a disc is less dependent on the precise details of the fit i.e. the assumed PSF, weights, *etc.*

It is clear from our analysis of single Sérsic and Sérsic+Exponential galaxies with BUDDA, GASP2D and PHI, that major disagreements appear when different images, weights and PSF values are used. In general, disc parameter estimates seem far more consistent between codes than bulge parameters to the code or the images used. However, bulge parameters are harder to derive accurately. Deviations between Gadotti (2008) and PHI could arise due to the different PSFs assumed. The reason for the linear correlation observed in the comparison is likely due to the limited resolution of the bulges. This would explain how we see large deviations for the bulge parameters and more of an agreement for the disc parameters.

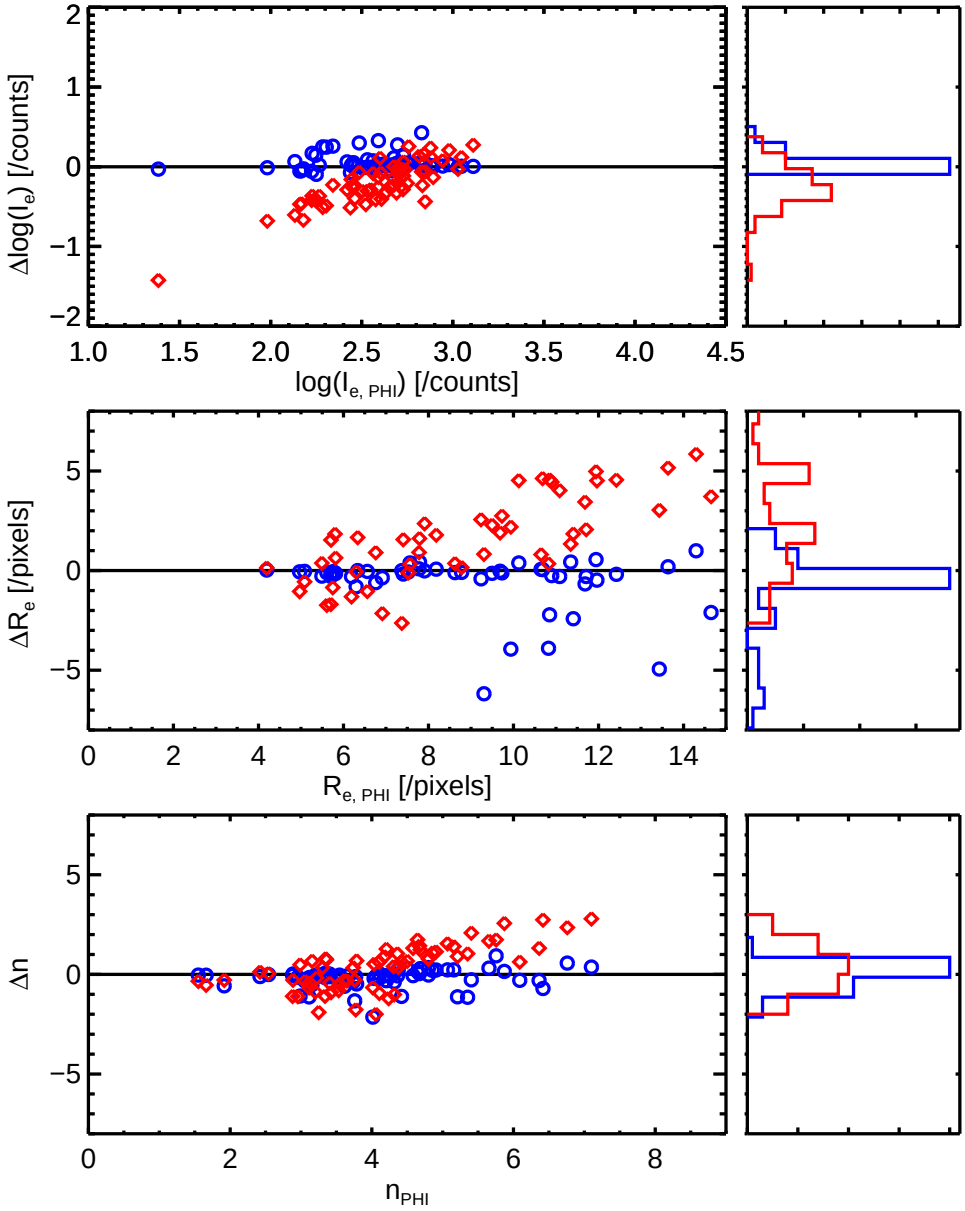


Figure 3.13: Differences between parameter estimates for the elliptical galaxies where the model is a single Sérsic profile. From the top to bottom the parameters are: the effective intensity, I_e, the effective radius, R_e and the Sérsic index, n. Blue open circles are the posterior medians given by PHI minus the results from GASP2D and the red diamonds are posterior medians given by PHI minus the results from Gadotti (2009). To the right of each panel is the distribution of the parameter residuals where the blue histogram is the difference between PHI and GASP2D and the red histogram is the difference between PHI and the results of Gadotti (2009).

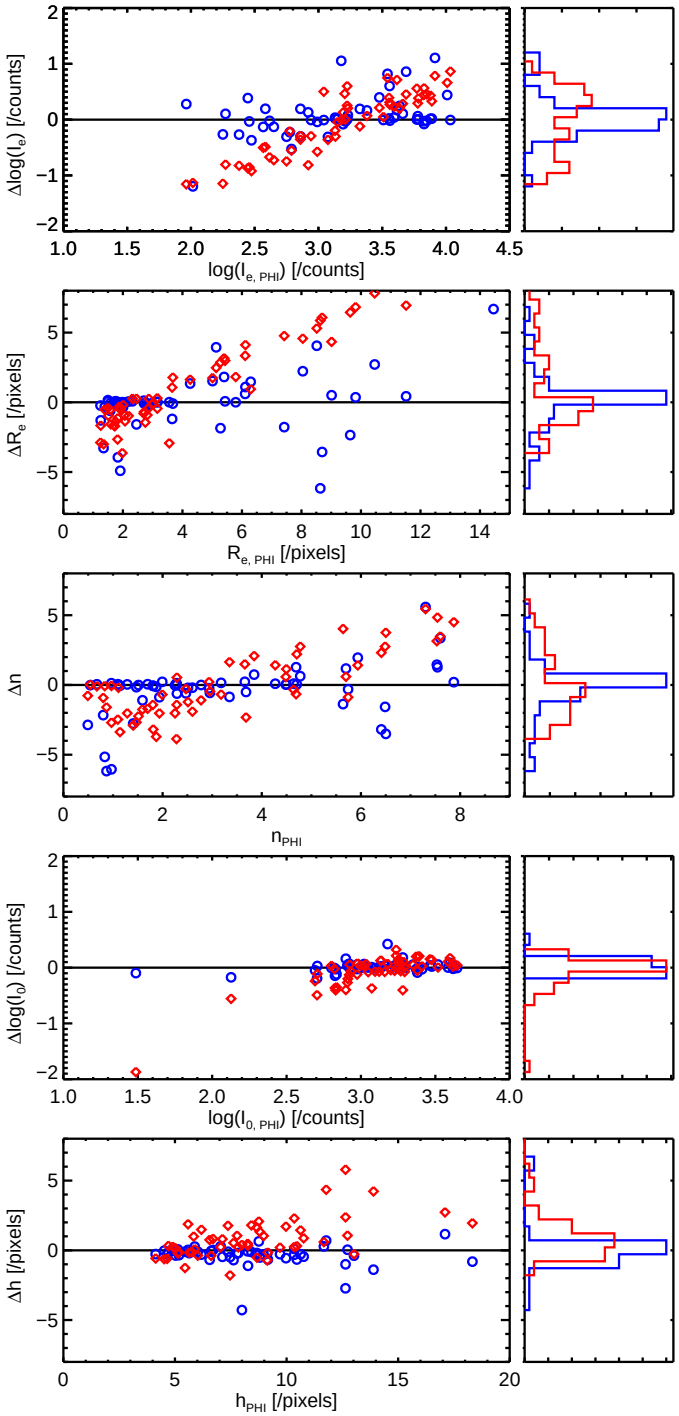


Figure 3.14: Differences between parameter estimates for the disc+bulge galaxies where the model is a Sérsic profile plus an exponential. From top- to-bottom the parameters are: the effective intensity, I_e, the effective radius, R_e , the Sérsic index, n, the central intensity I_0, and the scale length,h. See figure 3.13 for more details.

3.5 Posterior predictive checks and model comparison

In the previous sections we have constructed models and computed the posterior distribution for the parameter matrix for these models as well as shown some practical tests that have helped us understand the methodology. However, when we use `PHI` on future samples we will want to visually and quantitatively check that the results are robust. One such test is to directly compare a synthetic galaxy generated with the parameters predicted by `PHI` to the real image. Any discrepancy could be due to a poorly chosen model or poorly calculated systematics. The basic technique for checking that a model is a good fit to the data is to draw random values from the posterior distribution, replicate the data and then compare this sub-sample to the observed data. Any differences could indicate possible failings of a model.

3.5.1 Residual plots

By collapsing the 2D image of both the real and model galaxies into a one-dimensional (1D) form using ellipse-averaging we can check that a particular model fits the data correctly in an effective graphical representation. Given the fitted parameters, θ, we can predict the morphology of the galaxy data and calculate a residual.

We generate our residual plots by drawing randomly from the parameter posterior distributions and plotting a realisation of the 1D surface brightness profiles and residuals. This way we include the posterior uncertainty in θ rather than just using the most likely values.

Figure 3.15 shows the 1D surface brightness profile from bulge+disc galaxy from the Gadotti (2009) sample. In figure 3.15 we see the image data (black dots) with the grey region designating the root-mean square of the pixel values. The green band represents an ensemble of model fits generated from drawing randomly from the posterior distribution then convolving the model galaxies with the PSF. The left panel shows the single Sérsic compoent fit and the right panel the two-component fit. The lower panels in figure 3.15 show the residuals between the data and the model galaxies. It can be seen that the one-component model represents the central region of the galaxy fairly well but then deviates at large radii. This suggests to us that the model needs an extra level of complication in order to better represent the data. In the right hand panel, the red band represents the possible profiles for the bulge component and the blue band is showing that for the disc. The residuals show a much improved fit at $3 < R < 8$. However, the model profile still deviates in the outer regions of the galaxy. The further addition of a truncated disc could solve this problem

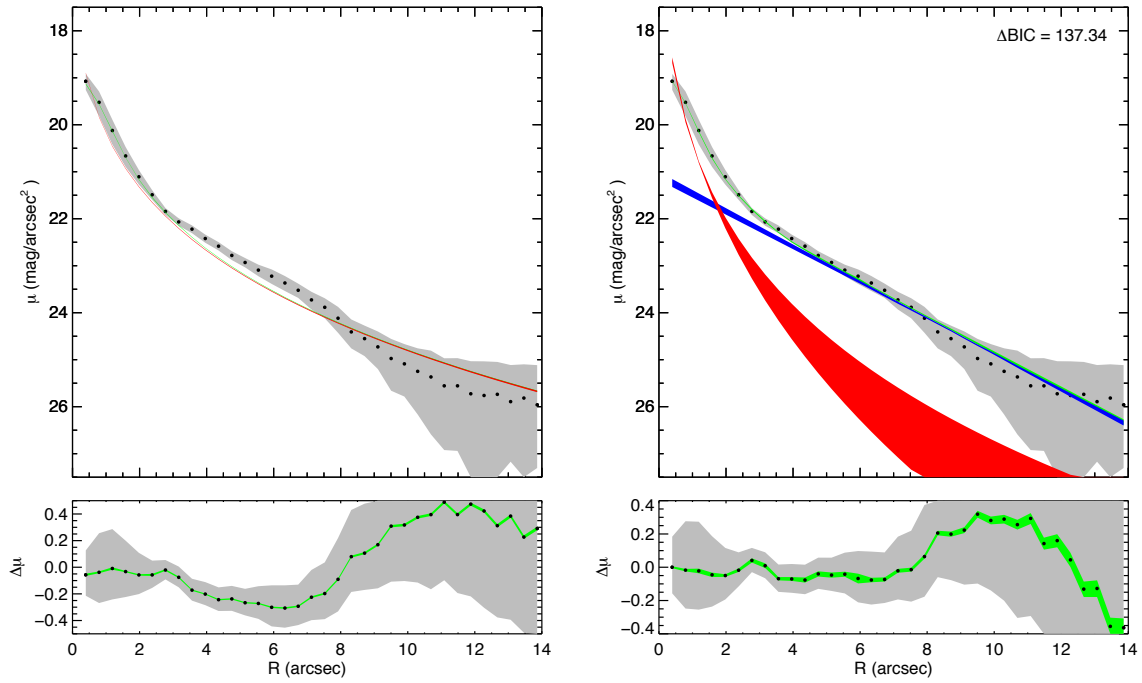


Figure 3.15: Ellipse-averaged surface brightness radial profile for an observed galaxy from SDSS (black dots) with the root mean square error from the pixel values in the image (grey region). The green band signifies model galaxies generated (and PSF convolved) from random draws from the output posterior distribution using a Sérsic and Sérsic+exponential model (left and right panels respectively). The blue band shows the random draws for the exponential component and the red band shows that for the Sérsic component without PSF convolution. The lower panel shows the residuals between the model (created with the posterior median) and the data.

(see Méndez-Abreu et al. 2017).

The graphical representation helps in determining which model best describes the data. In the next section we explore a more quantitative ways of finding the most probable model.

3.5.2 Results for the Bayesian model comparison

Figure 3.16 shows the ΔBIC values for all the synthetic galaxies (top panel) and SDSS galaxies (middle and bottom panels). We define the ΔBIC as the difference between the two-component fit and the one component fit i.e. $\Delta\mathrm{BIC} = \mathrm{BIC}_{Sérsic} - \mathrm{BIC}_{Sérsic+exponential}$. For every object we have fit a single Sérsic and a Sérsic +exponential model in order to calculate the ΔBIC. The single component/elliptical synthetic galaxies have ΔBIC values centred tightly on zero whereas the two-component/bulge+disc galaxies extend to $\Delta\mathrm{BIC} > 0$. The results from the synthetic galaxies suggest

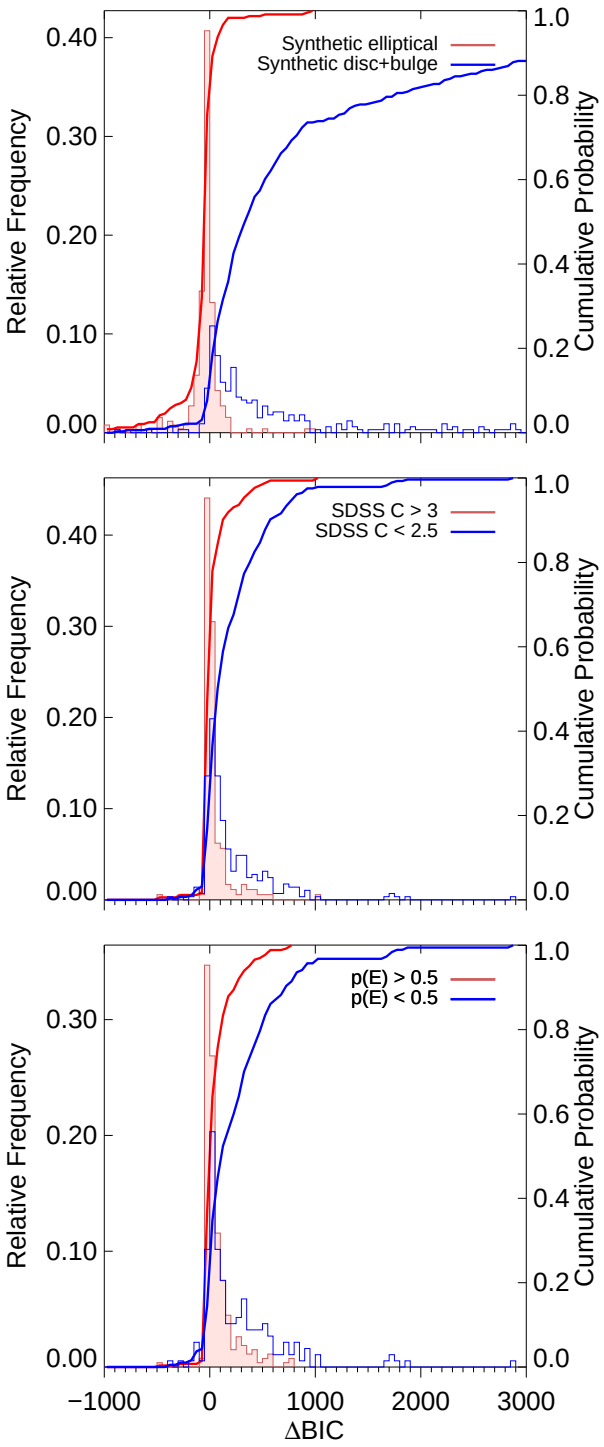


Figure 3.16: Histograms (and cumulative distributions) showing the ΔBIC distributions for synthetic single and two-component galaxies, the sample of SDSS galaxies classified by Gadotti (2009), and galaxies that are more likely to be single component according to the results from the machine learning approach by Huertas-Company et al 2011 (Top, middle and bottom panels respectively). As $\Delta\mathrm{BIC} = \mathrm{BIC}_{Sérsic} - \mathrm{BIC}_{Sérsic+exponential}$ a more positive value signifies the preferred model is that of two-components.

Table 3.3: Table showing the statistics for Figure 3.16 of the ΔBIC distributions in the synthetic and SDSS samples.

Synthetic ΔBIC				
	Mean	Median	16%	84%
Elliptical	-427.47	-31.09	-149.74	13.36
Bulge + disc	1218.51	341.64	35.13	2272.50
SDSS ΔBIC				
	Mean	Median	16%	84%
C >3	-89.70	2.78	-25.30	89.63
C <2.5	225.67	99.43	-4.29	480.77
p(E) >0.5	-16.88	18.82	-21.91	152.45
p(E) <0.5	280.88	134.56	3.69	551.71

that it is possible to classify a galaxy purely using the ΔBIC value. The results from the sample of SDSS galaxies are less clean cut. The bulge+disc galaxies defined by C have generally large positive values of ΔBIC with a median value of ΔBIC = 123.7 while classified as ellipticals have a median value of ΔBIC = 6.6. However, there is a tail of elliptical galaxies with large ΔBIC, and a large fraction of bulge+disc galaxies with ΔBIC < 0.

The plots in Figure 3.16 can be thought of as a comparison between classification schemes (Table 3.3 summaries the findings). The galaxies were previous classified in (Gadotti, 2008) using the concentration parameter, C (with ellipticals being defined with $C > 3$ and disc-dominated galaxies having $C < 2.5$). To further clarify the validity of the ΔBIC term we will can compare to the results from a machine learning. In Huertas-Company et al. (2011) a sample of SDSS galaxies were morphologically classified based on support vector machines. This approach gave a probability to each galaxy of being either an elliptical, S0, SAB, and SCD type morphology. The algorithms were trained on visual classifications from Galaxy Zoo first release catalogue. In the bottom panel of figure 3.16, we look at the distributions of the ΔBIC terms for classifying elliptical galaxies as those having a probability $p(E)$ greater than 0.5. The cumulative distributions are similar between the machine learning classifications and C.

In our synthetic galaxy population the ΔBIC term can be clearly used as a classification method to separate one- and two- component galaxies. However, in reality the complex structures of galaxies blurs the clear separation in ΔBIC as can be see in the bottom two panels in Figure 3.16. Even allowing for this, galaxies with higher values of ΔBIC will have a higher probability of being two-component galaxies. In future studies of galaxy structures the ΔBIC terms could be used alongside other classification methods.

3.6 Summary

We have presented a new synthetic galaxy imager algorithm that can be used in conjunction with fitting procedures for 2D photometric decompositions of galaxies. Synthetic stellar populations have been used to allow for a realistic modelling of the galaxies. Four different star formation histories were used to calculate the apparent magnitude in specified band-passes. The imager also uses a range of radial profiles to estimate the intensity as a function of radius received from an emitting source along with surface-brightness dimming laws for creating realistic galaxies at high redshifts. Furthermore, the images simulate realistic observations by convolving the image with a PSF that takes three different forms, as well as simulating the background noise levels.

We use a sample of synthetic galaxies to investigate the effect of observational attributes along with any internal code effects that could bias the end outputs. We find that the Sérsic index is effected by most parameters, most significantly the effective radius. With decreasing R_e the IQR for the fractional error on n increased (see Fig. 3.8). A similar increase in the IQR happened for the ratio between R_e and the PSF FWHM.

In this chapter we have also applied PHI to a sample of SDSS galaxies. The aim of this was to determine how consistent the new algorithm was compared to previous algorithms in obtaining parameter estimates. The first step showed that under the same image conditions i.e. the same galaxies with the same systematics, PHI and GASP2D achieved consistent results with minor scatter. This validates both algorithms and approaches when assessing galaxy structures in the nearby universe. We next compared the results of PHI with the results from the previous work of Gadotti (2009). The parameter values differed significantly however we believe this is due to different image conditions, such as the assumed PSFs and data releases.

We have also outlined our model comparison methodology that helps us to determine the most probable morphology for a galaxy. This involves using Bayesian residual plots for visual inspection in conjunction with a numerical model comparison values (ΔBIC).

4

Hierarchical Bayesian Modelling of galaxy scaling relations for nearby galaxies

4.1 Introduction

Observed physical properties of galaxies are not randomly distributed but instead follow trends commonly described by power laws. These trends form what is known as a galaxy scaling relation. Understanding the origin and nature of galaxy scaling relations is a fundamental quest of any successful theory of galaxy formation. Theoretical simulations of how a galaxy might form and evolve will be scrutinised, if they do not predict observed scaling relations robustly. Both broad classes of galaxy (i.e. elliptical and spiral) can be described by scaling relations and these have been compared to and used to calibrate numerical models. Some observed scaling relations can be reproduced to fairly good accuracy by invoking galaxy formation models that include virial equilibrium after dissipational collapse of spherical cold dark matter haloes and angular momentum conservation (e.g. Mo, Mao, & White 1998; Navarro & Steinmetz 2000; Firmani & Avila-Reese 2000; Shen, Mo, Shu 2002).

Scaling relations have been used for theoretical modelling of elliptical galaxies (Zeeuw & Franx 1991) and for investigating the formation of galactic discs in spiral galaxies (Dalcanton, Spergel & Summers 1997; Firmani & Avila-Rees 2000; Tonini et al. 2016). Two such scaling relations are between the brightness and the size parameters of the objects (Courteau, de Jong & Broeils 1996), and between the luminosity of the bulge to the total luminosity of the galaxy (Carollo et al. 2007). This is one of the fundamental relations that has been associated to the formation of elliptical galaxies. It has been observed that elliptical galaxies obey a tight scaling relation in a three-dimensional plane between central velocity dispersion, effective radius and the average surface brightness within one effective radius, which has been titled the Fundamental Plane. (FP; Djorgovski & Davis 1987; Dressler et al. 1987). The observed thinness of the FP indicates that the mass-to-light ratio follows a power-law without much scatter, putting tight constraints on models for the formation and evolution of elliptical galaxies. Before it was known that elliptical galaxies obey a FP relation, relations between parameter pairs in the FP were observed first, and are now thought of as projections of the FP. One such relation is the Kormandy Relation (Kormendy 1977, hereafter KR). It shows that the effective radius is anti-correlated with the mean effective surface brightness, and has been used to support a merger built scenario for ellipticals (Bernardi et al., 2003) as well as to demonstrate links between bulges and ellipticals (Bender et al. 1992).

All the studies so far have relied on a classical statistical approach to obtain the parameters. We are now in a position to take the Bayesian framework for the structure of galaxies to its next logical step, constructing global scaling relations using properly obtained posterior probabilities. This chapter focuses on how we can construct a new way to interpret scaling relations of galaxies using Hierarchical Bayesian Models (HBM).

We first provide an outline of what a HBM is and then provide examples of how we can obtain an estimate for the global parameter space using a piece-wise constant representation (Section 4.2). We then show the results for the population of galaxies classified into two groups (Section 4.4): either single or two-component systems based on the model prediction method described in the previous chapter. Finally in Section 4.5 we investigate how the results can change depending on how you classify a galaxy.

4.2 Hierarchical Bayesian models

Thus far, we have dealt with probabilistic models that have a set of parameters θ, which determine a probability distribution that describes a set of observations D. Using Bayes' Theorem, probabilis-

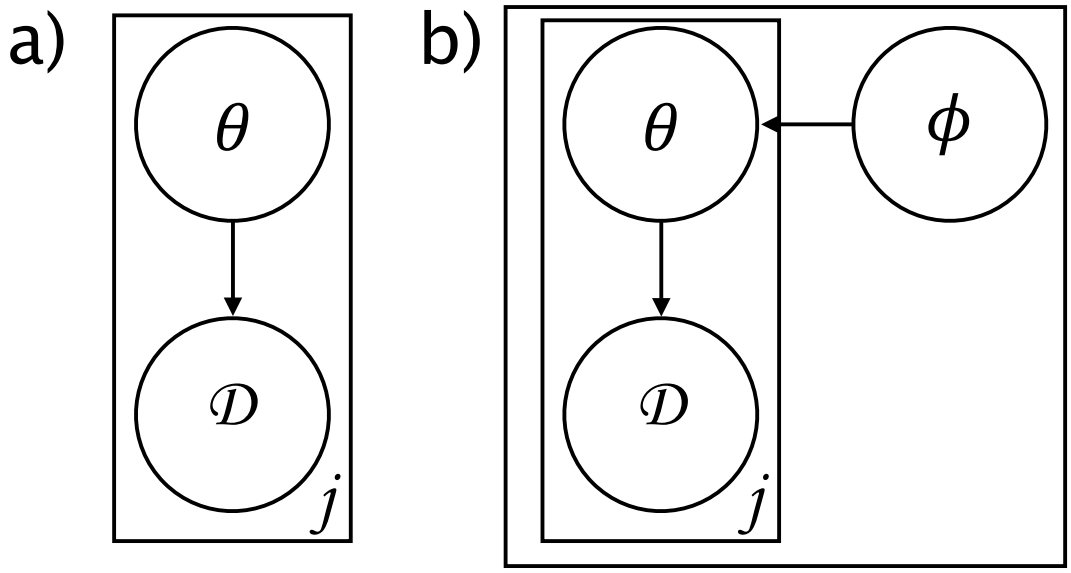


Figure 4.1: A succinct representation of a) a non-hierarchical model, where D is the jth observed data which is described by a set of parameters θ and b) a hierarchical model with D and θ taking on the same meaning but now θ is described by a set of hyperparameters ϕ.

tic estimates for the unknown parameters of individual galaxies can be computed coherently and consistently. The distribution precisely captures our uncertainty about the parameter estimate. As we have seen, this regular Bayesian model has the form $p(\theta|D) \propto p(D|\theta)p(\theta)$, i.e., the posterior probability is proportional to the product of the likelihood and the parameter priors.

In many situations, the parameters of a model have meaningful dependencies on each other. If the probability of one parameter can be conceived to depend on the value of another parameter, the model is said to be a *hierarchical Bayesian model* (HBM). Expressed formally, suppose again we have our observed data for individual galaxies D_g (where g indexes individual galaxies) which have been described by a model with a set of parameters θ_g. Given the known existence of scaling relations, it is reasonable to expect that estimates of the θ_g's for individual galaxies are related to each other. We can achieve an understanding of this assumption by using a prior distribution in which θ_g for each galaxy are viewed as a sample from a common *population distribution*. The problem can be approached hierarchically, with observable outcomes modelled conditionally on specific parameters, which are themselves given a probabilistic distribution modelled using higher level parameters, known as *hyperparameters*, ϕ. Previously, we assumed ϕ was known but now we include the uncertainty of ϕ in the model. The hierarchical aspect of the model means that ϕ will have its own prior distribution, $p(\phi)$. The joint posterior distribution between θ and ϕ is

$$p(\phi, \theta) = p(\theta|\phi)p(\phi), \tag{4.1}$$

where $p(\theta|\phi)$ is the conditional distribution of θ given ϕ and the joint posterior distribution can be expanded as

$$\begin{aligned} p(\phi, \theta|D) &\propto p(D|\phi, \theta)p(\phi, \theta) \\ &= p(D|\theta)p(\theta|\phi)p(\phi), \end{aligned} \tag{4.2}$$

with the assumption that $p(D|\phi, \theta)$ depends only on θ as ϕ only affects D through θ.

We must first assign a prior distribution ($p(\phi)$, or population model) to ϕ, to obtain the joint probability distribution $p(D|\phi, \theta)$. From a logical stand point we will assign a non-informative prior distribution to ϕ initially.

The following strategy is useful for drawing proposals from the joint posterior distribution $p(\phi, \theta|D)$, for simple hierarchical models.

1. Draw the vector of hyperparameters, ϕ, from their prior distributions, $p(\phi)$.
2. Draw the parameter vector θ from their conditional posterior distributions $p(\theta|\phi)$ given the drawn values of ϕ.
3. Draw predictive values D' from the posterior predictive distribution given the drawn θ.

The above steps are performed L times in order to obtain a set of L draws. From the joint posterior simulations of θ and D', we can compute the posterior distribution of any predictive parameter population of interest.

4.3 Piecewise constant representation population model

Let us consider a catalogue of galaxies with the intrinsic structural properties of the Sérsic profile (or a Sérsic + exponential profile) with the parameters: effective surface brightness μ_e, effective radius R_e, and Sérsic index n, (exponential parameters are: central surface brightness I_0 and the scale length h) in some reference photometric band. The parameters that describe the structure of a galaxy will be described by the parameter vector θ. The properties of these galaxies are assumed to be drawn from a unit-normalised distribution $p(\theta|\text{galaxy}, \text{survey})$, where the form of the distribution is explicitly dependent on the abundance of the sources as a function of the intrinsic properties and

Table 4.1: Summary of our notation

θ	Parameter vector
$\{\theta\}$	Set of intrinsic parameters
g	Galaxy index
N_{gal}	Total number of galaxies
N_{bin}	Number of bins for given parameter
F	Fractional galaxy count in bin
$\{F\}$	Set of all fractional bin counts F, summing to 1
ξ_h	Boundary to the hth bin
b_h	Galaxy count in the hth bin
$\hat{I}_g$	Observed photometric intensity for gth galaxy
$\{\hat{I}_g\}$	Observed photometric intensities for all N_{gal}

the characteristics and selection effects of the survey under consideration. For the remainder of this chapter we use the notation $p(\theta|\text{galaxy})$ to describe the population distribution of the galaxies but it should be noted that in reality it does have a dependence on specific survey effects.

For this study we model the galaxy population as a piecewise constant representation of $p(\theta|\text{galaxy})$. For this introduction into the modelling of the population we will consider only a single Sérsic profile model for individual galaxies and will expand to multiple components later in this section. This piecewise representation is parametrised by a set of coefficients $\{F_{i,j,k}\}$, such that:

$$\begin{aligned} p(\mu_e, R_e, n|\{F_{ijk}\}) = \sum_{i,j,k} \frac{F_{ijk}}{(\mu_{e_{i,max}} - \mu_{e_{i,min}})(R_{e_{j,max}} - R_{e_{j,min}})(n_{k,max} - n_{k,min})} \\ \times\Theta(\mu_e - \mu_{e_{i,min}})\Theta(\mu_{e_{i,max}} - \mu_e) \\ \times\Theta(R_e - R_{e_{j,min}})\Theta(R_{e_{j,max}} - R_e) \\ \times\Theta(n - n_{k,min})\Theta(n_{k,max} - n), \end{aligned} \tag{4.3}$$

where Θ is the Heaviside step function which is a mathematical function also known as the unit step function. The Heaviside step function is given by

$$\Theta(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{2} & x = 0 \\ 1 & x > 0 \end{cases} \tag{4.4}$$

Equation 4.3 corresponds to a mathematical description of a 3-dimensional histogram, where the probability of finding an object in the bin labelled ijk is F_{ijk}. The μ_e, R_e, and n are labelled with

$i = 1, ..., N_{\mu_e}$,$j = 1, ..., N_{R_e}$, and $k = 1, ..., N_n$, respectively, where N_{μ_e},N_{R_e}, and N_n are the total number of bins for each parameter respectively. The μ_e, R_e, and n all have limits as $(\mu_{e_{i,min}}, \mu_{e_{i,max}})$, $(R_{e_{j,min}}, R_{e_{j,max}})$ and $(n_{k,min}, n_{k,max})$. Here we adopt equal-sized contiguous bins, however this need not be the case. The 3D histogram provides a piecewise constant approximation of $p(\theta|\text{galaxy})$.

As was previously said, the piecewise representation is a type of interpretation which can be classified under the name *Bayesian histograms* where the goal is to obtain a Bayes estimate of the density $p(\theta|D_g)$ (we use D_g to signify an individual galaxy data). The histogram is often used as a simple form of density estimate. We now turn our attention to describing a more general implementation of this.

We have adopted a flexible parametric version of the histogram that helps to motivate the fully non-parametric Bayesian density estimation for the model which describes our population, similar to Leistedt et al. (2016). We initially have pre-specified knots (or bin boundaries) $\xi = (\xi_0, \xi_1, ..., \xi_{N_{bin}})$ to define our histogram estimate, with $\xi_0 < \xi_1 < \cdots < \xi_{N_{bin}}$ and $\theta_i \in [\xi_0, \xi_{N_{bin}}]$, thus a probability model for the density that is analogous to the histogram is as follows:

$$p(\theta|D_g) = \sum_{h=1}^{N_{bin}} \Theta_{\xi_{h-1}<\theta\leq\xi_h} \frac{F_h}{(\xi_h - \xi_{h-1})}, \tag{4.5}$$

where F is an unknown probability for each bin and where $\Theta(\cdot)$ is the Heaviside step function as above. Bayes theorem additionally requires a prior distribution for the probabilities $\{F\}$, which we take to be a Dirichlet$(a_1, ..., a_k)$ distribution:

$$p(F|a) = \frac{\Gamma\left(\sum_{h=1}^{N_{bin}} a_h\right)}{\prod_{h=1}^{N_{bin}} \Gamma(a_h)} \prod_{h=1}^{N_{bin}} F_h^{a_h - 1} \tag{4.6}$$

where $\Gamma(\cdot)$ is the gamma function. We use a uniform (maximally uninformative) prior on F,

$$p(F) = (N_{\text{bin}} - 1)! \delta_D\left(1 - \sum_{h=1}^{N_{\text{bin}}} F_h\right) \prod_{h=1}^{N_{bin}} \Theta(F_h), \tag{4.7}$$

where $\delta_D(x)$ is the Dirac delta function. In the case where the parameters of individual galaxies are known thus enabling a noiseless analysis, the numbers in each bin can be calculated by

$$b_h = \sum_{g=1}^{N_{\text{gal}}} \Theta(\theta_g - \xi_{h-1}) \Theta(\xi_h - \theta_g) \tag{4.8}$$

The likelihood of the binned data is given by a multinomial distribution

$$p(\{b_h\}|\{F_h\}) = N_{\rm gal}! \prod_{h=1}^{N_{\rm bin}} \frac{F_h^{b_h}}{b_h!}. \tag{4.9}$$

Combining the above prior and likelihood leads to a posterior for the Dirichlet distribution of

$$p(\{F_h\}|b_h) = N_{\rm gal}!(N_{bin}-1)!\delta_D\left(1-\sum_{h=1} F_h\right)\prod_{h=1}^{N_{\rm bin}} \frac{\Theta(F_h)F_h^{b_h}}{b_h!}. \tag{4.10}$$

Thus inferring the coefficients $\{F_h\}$ when the θ's are known is possible thanks to this analytic posterior that only requires the bin counts $\{b_h\}$.

The Bayesian histogram estimator does an adequate job approximating the true density, but the results are sensitive to the number and locations of knots. The Dirichlet prior distribution is perhaps not the best choice due to the lack of sampling across bins, but it does have the advantage of conjugacy[1] and simplicity in interpretation.

However, in reality the θ's for the individual galaxies are unknown and thus must be included in the final posterior probability. These must be inferred simultaneously with the underlying distributions using the photometric observations of the galaxies within the population.

In the imaging of a galaxy, the main observable is the photon intensity I, measured in some photometric band. These photon counts eventually become the images we receive and for a given galaxy g, we have noisy measurements of its photometric intensity $\hat{I}_g$ with errors $\sigma_{\hat{I}_g}$ within an image with a total number of pixels N_p. We define as in previous chapters a multidimensional Gaussian likelihood function in intensity space,

$$p(I_g|\theta_g) = \mathcal{N}\big(I(\theta_g); \hat{I}_g, \sigma^2_{\hat{I}_g}\big) \tag{4.11}$$

where $\mathcal{N}$ stands for a normal (i.e. Gaussian) distribution. The population model and the intensity likelihood can be combined into the hierarchical model. The observed quantities of this model are the set of $N_{\rm gal}$ intensities, denoted by $\{\hat{I}_g\}$. The parameters of interest are the intrinsic parameters $\{\theta\}$ as well as the population parameters $\{F_h\}$. The full, joint posterior distribution of these parameters reads

[1] If the posterior distributions are in the same family as the prior probability distribution (*for example* they are both Gaussian) both the prior and the posterior are said to be conjugate.

$$p(\{\theta\},\{F_h\}|\{\hat{I}_g\}) \propto p(\{F_h\}) \prod_{g=1}^{N_{\rm gal}} p(\{\hat{I}\}_g|\theta_g) p(\theta_g|\{F_h\}) \tag{4.12}$$

This implied posterior is the expanded form of the Bayes equation when we include the hyperparameters (see Section 4.2) and it does not have an analytic form. However, it is possible to directly draw samples from this posterior distribution using a two step Gibbs sampler, because the conditional posterior distributions can be easily sampled. The Gibbs process goes as follows:

1. A sample of $\{F_h\}$ is drawn from $p(\{F_h\}|\theta_g,\{\hat{I}_g\})$. This follows the Dirichlet model of Eq. 4.10 where the number of counts b_h are calculated from the $\{\theta_g\}$ of the previous Gibbs iteration or some initial uniform guess.

2. $\{\theta_g\}$ are then updated using the newly drawn $\{F_h\}$ by looping over galaxies and updating $\{\theta_g\}$ using
$$p(\theta_g|\{F_h\},\{\hat{I}\}_g) = \sum_{h=1}^{N_{bin}} \frac{F_h \times p(\{\hat{I}\}_g|\theta_g)}{(\xi_h - \xi_{h-1})} \times \Theta(\theta_g - \xi_{h-1})\Theta(\xi_h - \theta_g) \tag{4.13}$$

As for a classical Gibbs sampler, alternatively drawing $\{F_h\}$ and θ_g from the previous conditional distributions allows one to explore the full posterior distribution. One downside to the two step Gibbs sampler for our population of galaxies is that in order to fit simultaneously the population and the individual galaxy parameters, large computational power is needed.

To bypass this problem we can resample from individual posterior distributions of $p(\theta_g|\{F_h\},\{\hat{I}\}_g)$ using the updates of $\{F_h\}$. So all we need are the individual posterior distributions for each galaxy which enables us to skip the second Gibbs sampler. We fit each galaxy individually with `PHI`. We then implement the following process for the HBM:

1. A sample of $\{F_h\}$ is drawn from $p(\{F_h\}|\theta_g,\{\hat{I}_g\})$. Similarly to what was mentioned above, we use the Dirichlet model of Eq. 4.10 where the number of counts b_h are calculated from the $\{\theta_g\}$. For the initial iteration we choose a random point from each of the galaxies posterior probabilities, $p(\{\hat{I}\}_g|\theta_g)$, assuming a flat probability over the entire parameter space.

2. As we have used uniform priors for the galaxy parameters we can select points from individual posterior distributions using the previous iterations $\{F_h\}$. $\{\theta_g\}$ is then updated using the newly drawn $\{F_h\}$ by looping over galaxies and updating $\{\theta_g\}$ using $p(\theta_g|\{F_h\},\{\hat{I}\}_g)$. This is implemented by weighting each point in the parameter distributions for each galaxy with $p(\theta_g|\{F_h\},\{\hat{I}\}_g)$ and then updating $\{\theta_g\}$ by randomly picking from the weighted distributions.

3. A sample of $\{F_h\}$ is again drawn from $p(\{F_h\}|\theta_g, \{\hat{I}_g\})$ using this newly updated $\{\theta_g\}$.

The above steps are repeated for 10^5 iterations with the first 1000 iteration being removed for the burn-in. This solution vastly improves the speed of the method but we must question of whether the HBM under a resampling process can recover an underlying population trend. We now show the results of the HBM on a toy sample to attempt to answer this question.

4.3.1 One-dimensional test case

To check the practical implementation of the hierarchical model using the Bayesian histograms we focus on an application in which we have a population with each object having a given parameter value x_i where i indexes an object within the group so the $X \in \{x_1, ..., x_N\}$. We implement the following hierarchical model:

$$\begin{aligned} x_i &\sim \mathcal{N}(\mu_i, \sigma_{x_i}) \\ X &\sim \mathcal{N}(\mu, \sigma) \end{aligned} \tag{4.14}$$

where μ_i and σ_{x_i} are the means and standard deviations of the normal distributions for individual objects which combine to form the global population of objects which also follow a normal distribution with parameters μ and σ. This toy model illustrates our preconceptions of the distribution that galaxy population parameters could potentially follow, for example a sample of elliptical galaxies may have their Sérsic indexes lie in a Gaussian distribution. Using this Bayesian histogram method allows for a flexible, unbiased population model. In future studies, if one would like to parametrise a given population, e.g. by a power law slope, the model can be justified by referring to the probabilistic structure derived by our Bayesian histogram representation.

We can now apply our hierarchical model to gain information about the population by utilising information on individual objects. We follow the Gibbs sampling process described previously to sample from the global population model which re-samples from individual galaxy posterior distributions to demonstrate the models ability to recover the probabilistic structure of the population.

Figure 4.2 shows the population distribution of X represented by the black solid line where the number of global points (i.e., the N_{gal} in Eq.4.12) is 300. These were generated from Eq. 4.14 with $\mu = 5$ and $\sigma = 2$ (black dashed line)with each x generated given a probability distribution function with $\sigma_{x_i} \sim U(0.01, 0.5)$ (black solid line). We also add Gaussian noise to make this toy model more realistic. The posterior distribution obtained with the Gibbs sampler (re-sampling method) are

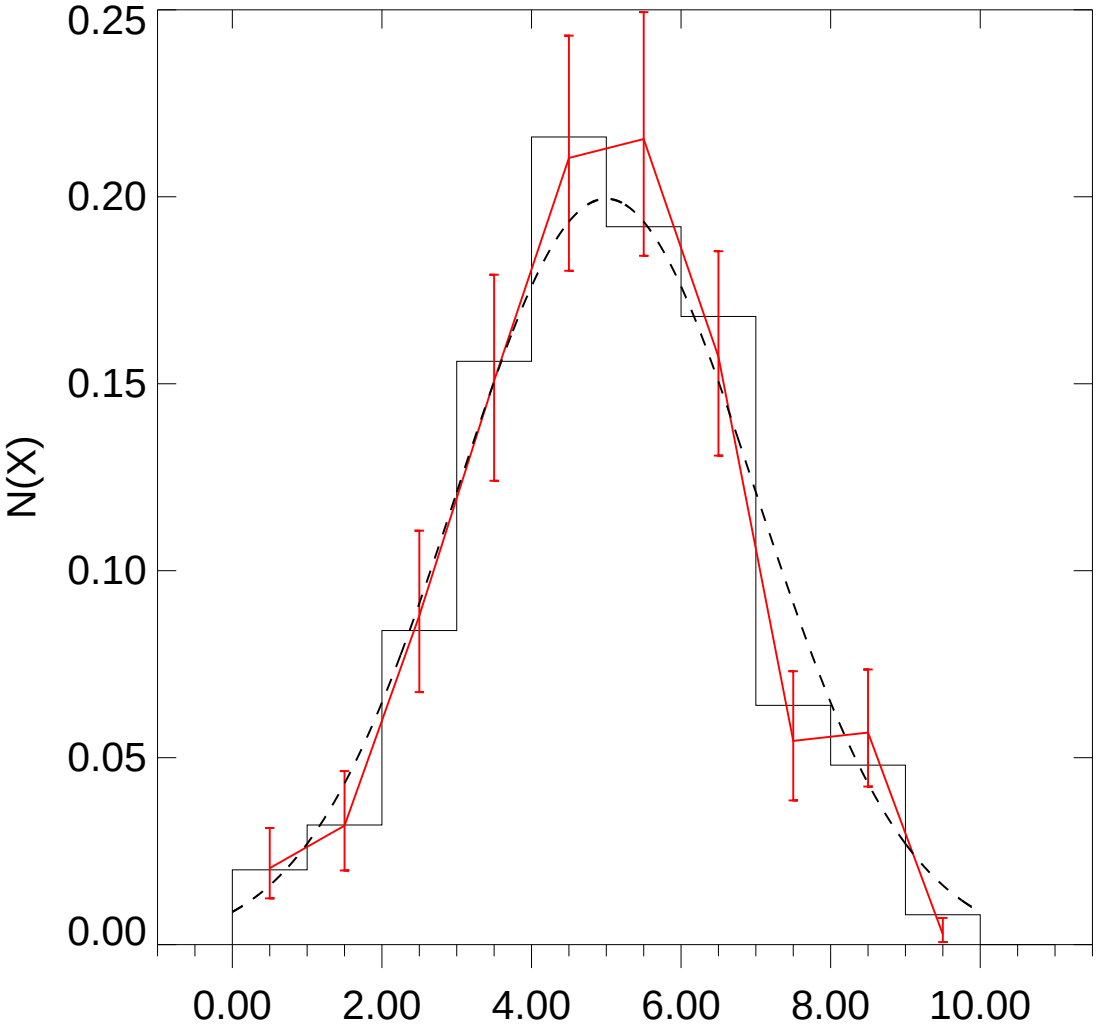


Figure 4.2: The global distribution of parameter X. The solid black line shows the input true distributions while the distributions obtained with the inferred parameters f_h are shown with the red solid line with the 16 and 84% percentiles shown in the error bars. The dashed black line shows the true distribution without any additional errors.

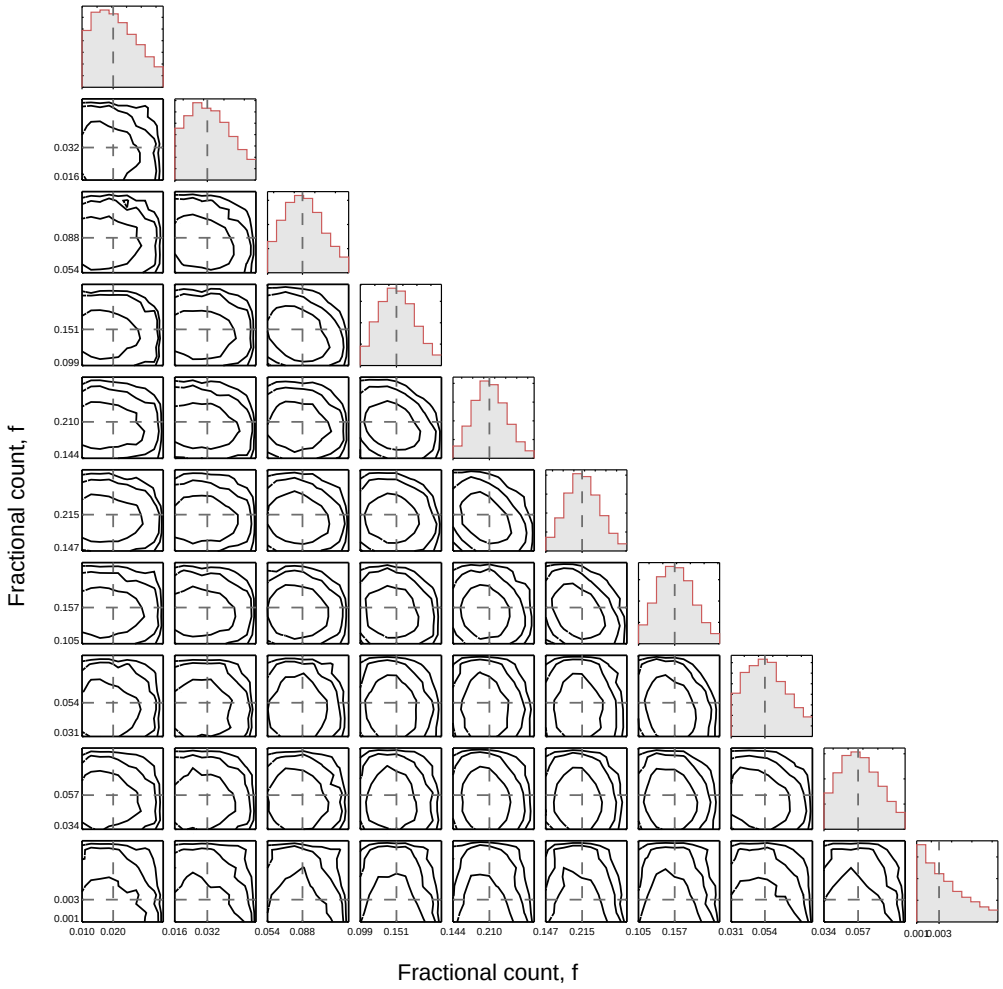


Figure 4.3: Posterior distributions of the fractional count parameters f for our 1D test case. Adjacent bins are anti-correlated as expected from the Dirichlet model, as adjacent bins will likely draw from similar galaxy posteriors. Dashed lines show the posterior medians for each fractional count parameter.

shown by the solid red line with their 16 and 84% percentiles. The distribution of X is successfully recovered with meaningful uncertainties. The posterior distribution also follows the underlying true distribution of this toy model (without the Gaussian noise) indicated by the dash line.

Figure 4.3 offers a closer look of the posterior distributions on the overall fractional counts for each of the 10 bins in our 1D test case. As expected, adjacent bins are anti-correlated, and correlation strength decreases with the distance between bins. Anti-correlation is a feature of the Dirichlet model but is enhanced due to adjacent bins drawing from bordering galaxy posterior probabilities.

4.3.2 Two-dimensional test case

We can extend our toy model to cover multiple dimensions (here we only show a two-dimensional (2D) plane). If we consider an overall population X (where $X \in \{x_1, x_2, ..., x_N\}$) that follows a linear regression so that

$$Y = \beta_0 + \beta_1 X(x) + \epsilon \tag{4.15}$$

where ϵ_i has the form $\epsilon \sim \mathcal{N}(0, \sigma)$ (σ is some scatter along the linear relation) and β_0 and β_1 are the linear coefficients, i.e., the intercept and the gradient. As for the 1D Gaussian toy population,

$$x_i \sim \mathcal{N}(X_i, \sigma_{x_i}), \tag{4.16}$$

where μ_i and σ_i are the means and standard deviations of the normal distributions for individual objects which combine to form the global population described by Eq.4.15. Again we want to test whether the Bayesian histogram can recover the hidden structure presented to us in some galaxy population parameters but this time in a multi-dimensional fashion. If one was to perform this type of analysis on a real sample of objects, and the results of the histogram suggested a linear relation then the logical step of fitting a hierarchical linear model would be justified. However, with that in mind we do not want to take away the power of having a truly non-biased probabilistic description of the population parameter space and its uses for future studies.

Figure 4.4 shows the population distribution of X and Y used for the test where the number of global points (i.e. the N_{gal} in Eq.4.12) is 300. These were generated from Eq. 4.15 with $\beta_0 = 0$, $\beta_1 = 2$ and $\sigma = 1$. Each $[x_i, y_i]$ point generated, has a probability distribution function with $\sigma_{x_i} \sim U(0.01, 0.5)$ and $\sigma_{y_i} \sim U(0.01, 0.5)$ with the covariance between them as $\sigma_{xy_i} \sim U(0.01, 0.5)$. We add Gaussian noise to make this toy model more realistic. Figure 4.5 shows the posterior medians for the Bayesian histogram description of this 2D plane. The distribution of X is successfully recovered with meaningful uncertainties. The posterior distribution also follows the underlying true distribution of this toy model (without the Gaussian noise) indicated by the solid line with the error 1-sigma uncertainty given to the population shown by the dashed line.

The top panel of Fig. 4.6 shows an example of the HBM in the $R_e - n$ plane for the sample of two-component SDSS galaxies. The bin values represent the median posterior probability for the

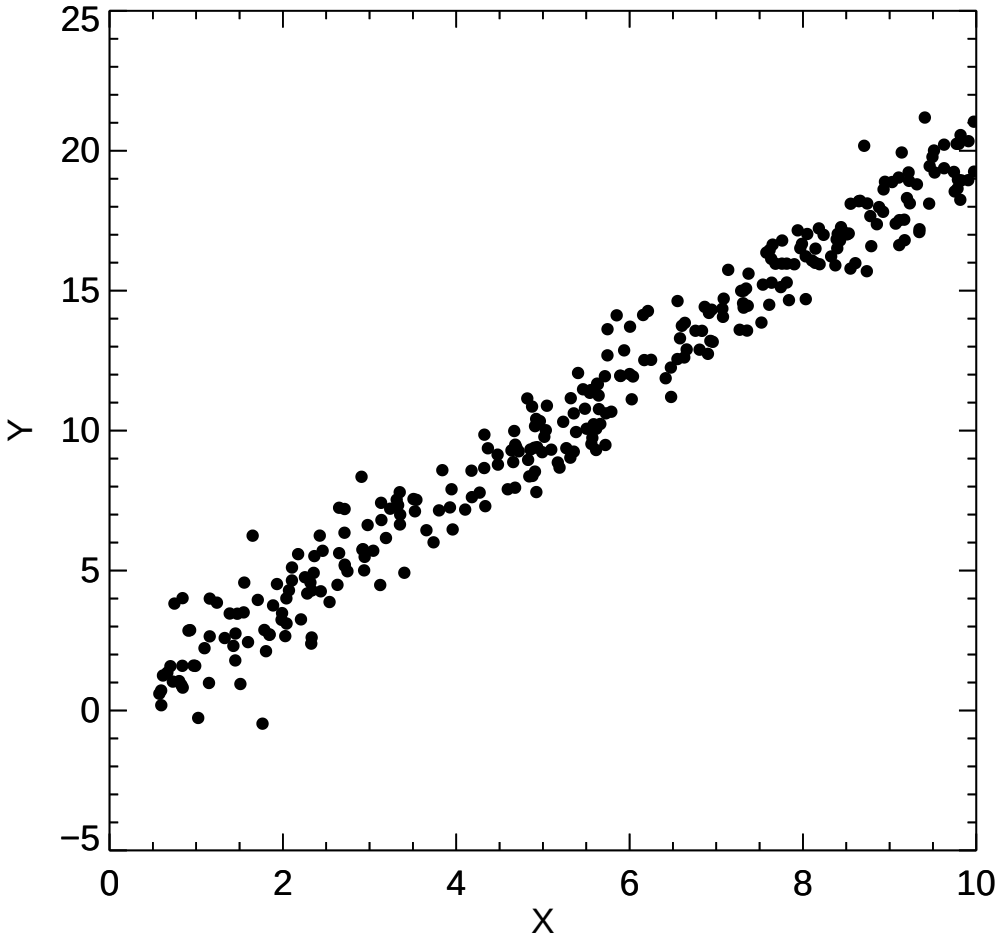


Figure 4.4: The global population used as a test case for the 2D Bayesian histogram. Each point has a full probability density function that is not shown.

$R_e - n$ plane. We can compare this to the middle panel of Fig. 4.6 where we have combined the probability distributions for the entire two-component population. The HBM contains information about the probabilities in the $R_e - n$ plane. The same can not be said for the second panel as distributions can overlap in the outer wings which could hide the true global trend. The bottom panel of Fig. 4.6 shows a sub-sample of 20 two-component galaxy distributions taken from the population that was used to make the middle panel (a sub-sample had to be taken for visual interpretations). Bayesian shrinkage helps the HBM show the true underlying population trends, whereas the stacking of distributions in the middle panel can bring about a biases in the interpretation.

4.4 Galaxy population trends in the structural parameters

Our approach benefits from a key property of hierarchical probabilistic models: *shrinkage* (Gelman et al. 2014). Not only does the hierarchical model allow us to estimate the underlying distribution

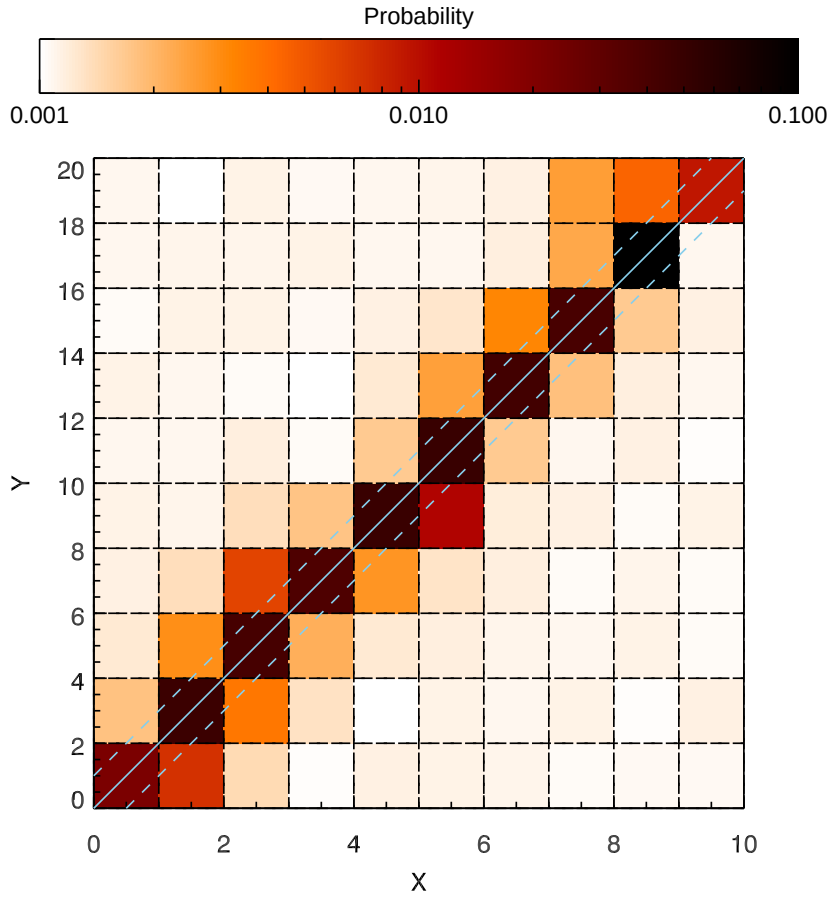


Figure 4.5: The 2D Bayesian histogram representation of the test population X and Y. The bin colours show the median proabability for the fractional count in each bin. The solid line shows the input model with the 1-sigma uncertainty added to the population shown by the dashed lines.

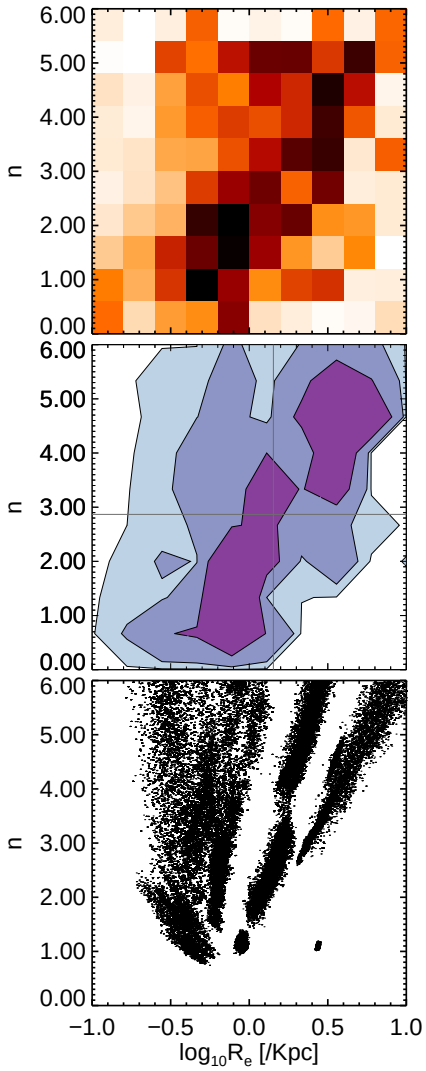


Figure 4.6: Figure showing the benefits of using the HBM. The top panel shows an example from the HBM of the $R_e - n$ plane for the two-component galaxies from SDSS (see Section 4.4.3). Refer to the colour bar in Fig. 4.5 for the bin values. The middle panel shows the 1, 2, and 3σ contours for the posterior distributions for the entire SDSS two-component population. The grey lines indicate the medians for the distributions. The bottom panel shows the full posterior probability distributions for a sub-sample from the two-component SDSS galaxies.

but it is also applied as a prior to shrink the uncertainties on the properties of individual objects. This Bayesian shrinkage naturally occurs within the HBM when computing or sampling from the full posterior distribution. In the case of galaxy structural parameters, previous works have shown that they follow scaling relations with some degree of scatter that can be intrinsic or due to the data itself. Our new HBM improves on this scattered distribution by taking into account the population trends and the individual galaxy measurement at the same time.

4.4.1 Data and sample classification

We use the sample of local galaxies from the previous chapter i.e. the SDSS sample from Gadotti (2009). These galaxies reside in the local Universe with a redshift range of $0.04 \leq z \leq 0.06$ and stellar masses greater than 10^{10} solar masses. The total number of galaxies is 655. Gadotti (2009) originally separated the sample based on their concentration parameter values (see Chapter 3), but we have shown that some of the galaxies classified as ellipticals had evidence of an outer disc component thus making them two-component systems by our definition. It is important that we distinguish between systems with one or two components as this will lead to more information to better understand galaxy evolution.

We have performed 2D photometric decomposition on all the galaxies in the sample with both a single and a two-component model (*i.e* a single Sérsic and a Sérsic + exponential model). Fittings that fail to converge (i.e. the simultaneous MCMC chains fail the final convergence tests discussed in Section 2.4.3) are removed from the final sample. There are 17 galaxies that fail to converge when fit with a bulge-disc model. After examining these failed galaxies we find that a single Sérisc with a $n \sim 1$ fits reasonably well. This suggests that these galaxies either have bulges that can not be constrained in the fitting process due to their size or that they are bulgeless, pure disc galaxies. After the fitting of each galaxy, we calculated the ΔBIC term. The BIC term is an estimation to the Bayes factor and was first introduced in Raftery (1995). In Raftery (1995) it is said that the size of the ΔBIC can determine the more preferred model, where a $\Delta\text{BIC} > 10$ roughly equates to a Bayes factor value > 150 thus there is strong evidence for one model over another. However, as part of our model building strategy we wanted to calibrate our ΔBIC values to compare to other classification methods. Huertas-Company et al (2011) used a machine learning approach to classify SDSS galaxies based on a Support Vectors Machine algorithm using non-parametric galaxy properties; colour, shape parameters, and concentration. This resulted in each object being given a probability of it being a specific Hubble morphology type. Our interpretation of the ΔBIC determines the most probable photometric model for a galaxy; either a one or two-component system. The machine learning

classification accounts for properties that have previously been used to separate ellipticals to spirals. The two methods do not provide exactly the same information as galaxies that would be considered pure discs from the machine learning method would be considered as one component in ours.

For our studies we can bridge the two classification methods using the cumulative distributions of our sample. In Figure 4.7 we compare the ΔBIC values vs. the probability of a galaxy being elliptical ($P(Elliptical) > 0.5$) presented in the results of Huertas-Company et al (2011). We then perform a Kolmogorov-Smirnov (KS) test on the two distributions at intervals of increasing ΔBIC i.e. a KS test is preformed on increasing iterations of $\Delta\text{BIC} = 10$ between the two distributions. The KS test can give us an indication where the two distributions diverge thus giving an idea of a suitable ΔBIC threshold value to describe a galaxies preferred model. The middle panel of Fig. 4.7 shows how the distance between the two cumulative distributions decreases as ΔBIC increases suggesting that at low values of ΔBIC the two distributions are similar but as ΔBIC becomes more positive the distributions become more distinct. The distance curve flattens after a $\Delta\text{BIC} \approx 25$. This supports the original grades of evidence given in Raftery (1995). For our classification scheme we thus use $\Delta\text{BIC} = 25$ to separate our one and two-component galaxies. In Section 4.5 we compare the two classification schemes to further justify our choice.

4.4.2 Selection of the global population model

Before we look at the results from the HBM we must first assess whether or not the inference methodology is indeed giving us a robust, accurate description of the population. As we have shown in the previous sections, the hierarchical model can be applied to multiple dimensions or just a single dimension; depending on the population trend we intend to study. However, as we include more parameter dimensions to our hierarchical model the number of hyperparameters, i.e., the number of bins, increases to the power of the number of parameters. This is because we used a piecewise constant representation where every bin of the histogram has an individual parameter. So for the one dimensional HBM we would have typically 10 bins per parameter so 10 hyperparameters, but for a five dimensional HBM this number would increase to 10^5 hyperparameters. As the number of galaxies in our sample is fixed this can become a problem, as we do not want to over-fit. For our situation we can address this drawback by focussing on a small number of dimensions thereby reducing the number of hyperparameters.

To ensure that reducing the number of dimensions does not diminish the reproducibility of the underlying population we have performed a test whereby we compare the HBM results of a run

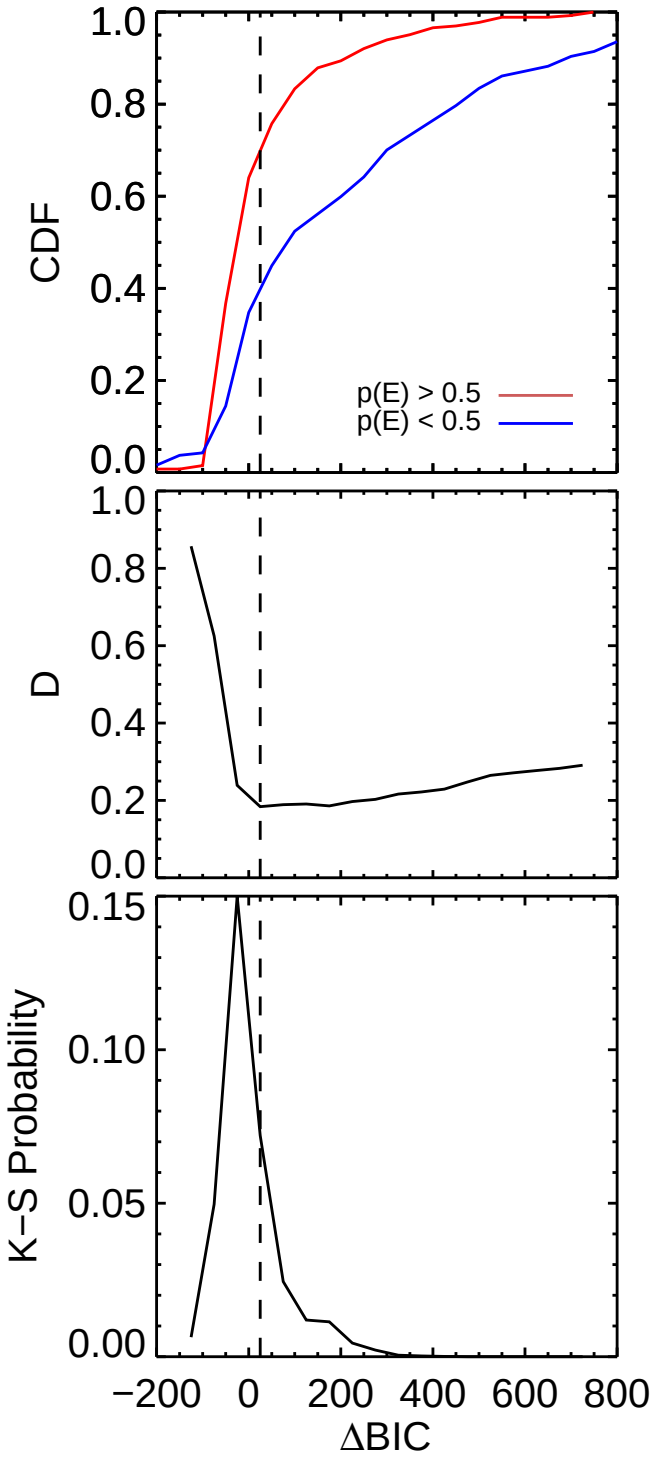


Figure 4.7: The top panel shows the cumulative distributions of the ΔBIC values for the sample of SDSS galaxies split by their likelihood of being elliptical as derived by Huertas-Company et al 2011. As ΔBIC increases the distance between the two distributions decreases with a minimum at ΔBIC ≈ 25 (As $\Delta\mathrm{BIC} = \mathrm{BIC}_{Sérsic} - \mathrm{BIC}_{Sérsic+exponential}$ a more positive value signifies the preferred model is that of two-components). The red line shows galaxies that have a probability of being elliptical greater than 0.5 and the blue line shows the remaining galaxies. The next two panels show the results of a Kolmogorov-Smirnov test on the two distributions.

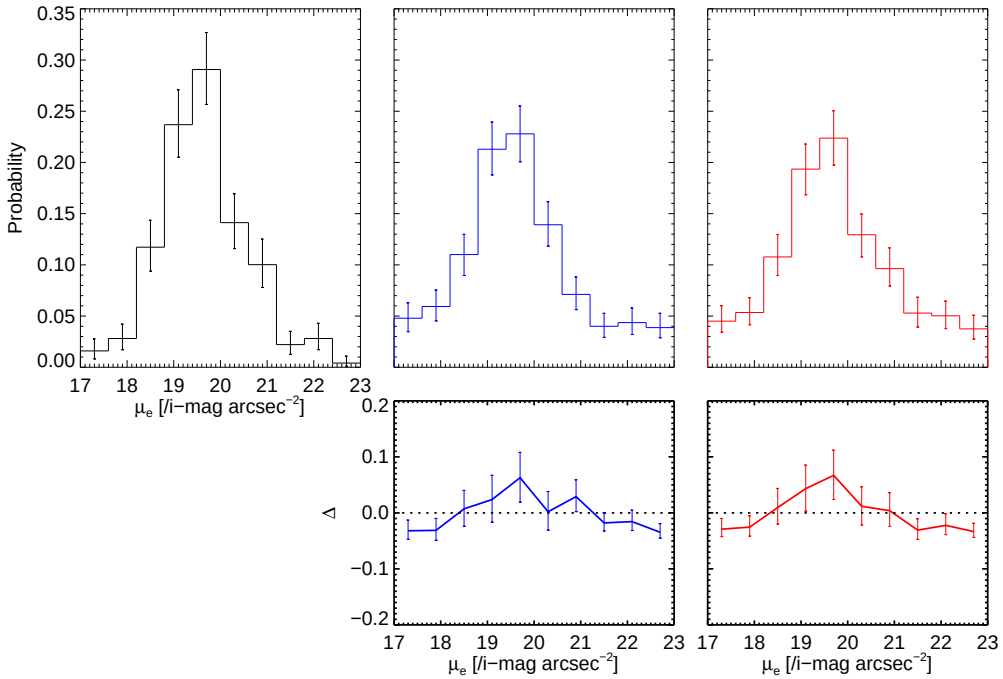


Figure 4.8: Results of the HBM for a sample of galaxies defined to be most likely single component system (ΔBIC < 25). The black histogram shows the results of the one dimensional HBM with error bars showing the 16% and 84% percentiles for the effective surface brightness. The blue histograms show the marginalised representation of the two dimensional HBM results between μ_e and R_e , whereas the red histogram shows the HBM results between μ_e and the Sérsic index, n. The bottom panels show the residuals between the 1DHBM and the 2DHBM in each case.

on a single dimension (hereafter 1DHBM) to that of two dimensions (hereafter 2DHBM). We first restricted our attention to the single component galaxies chosen using the ΔBIC (i.e., galaxies that are more likely to be described by a single Sérsic profile, ΔBIC < 25). The top left panel of Figure 4.8 shows the results of the 1DHBM on the effective surface brightness,μ_e, of the single component galaxies. In the next two panels we show the marginalised one-dimensional representation of the 2DHBM for the global population of μ_e verses the effective radius, R_e, and the Sérsic index, n, the blue and red histograms respectively. The lower plots show the difference between the 1DHBM and the 2DBHM for μ_e. It becomes obvious here that as the data is spread over more hyperparameters the more the priors speak. So in the presence of little data, the prior distributions would have more influence, and consequently the inference can be conditioned by the choice of prior. For the HBM we have chosen uniform priors so as we increase the number of dimensions, the bin fractions begin to flatten, until eventually they are dominated by the prior. For the step from a 1DHBM to a 2DHBM this effect brings the wings of the distribution up and the peak down. Similar results are also seen in Figure 4.9 where we show the 1DHBM and the 2DHBM for the results of R_e and n, the black histograms showing the 1DHBM and the blue histogram showing the the marginalised representation of the 2DHBM between the two Sérsic parameters. Again it can be seen that including an extra population parameter into the problem, results in a flatter distribution. For the remainder of this chapter we have focused on the interpretation of the underlying population trends through two dimensional plains and one dimensional distributions of the population parameters as a full multi-parameter fit would not be possible.

4.4.3 Population trends in the Sérsic parameters of one- and two-component galaxies

Figure 4.10 shows the results obtained after performing the 1DHBM on the sample of galaxies that are more likely to be two-component systems (ΔBIC < 25). There are hints of a bimodality in the R_e, n, and B/T population distributions. This bimodality has been used to separate the bulge population in to two groups: classical and disc like bulges. A common practice is to use the Sérsic index as a classification threshold of classic to disc-like bulges. Usually this threshold value is taken to be $n = 2$ (Ho & Kim 2014; Ribeiro et al. 2016). This classification scheme has been shown to be erroneous and misleading in the majority of cases (Gadotti 2009). For our analysis, we instead separate galaxies only into one- and two-component systems.

Comparing the 1DHBM Sérsic profile parameters for both the single component and the two-component galaxies we find that the Sérsic profiles of the single component galaxies are more ex-

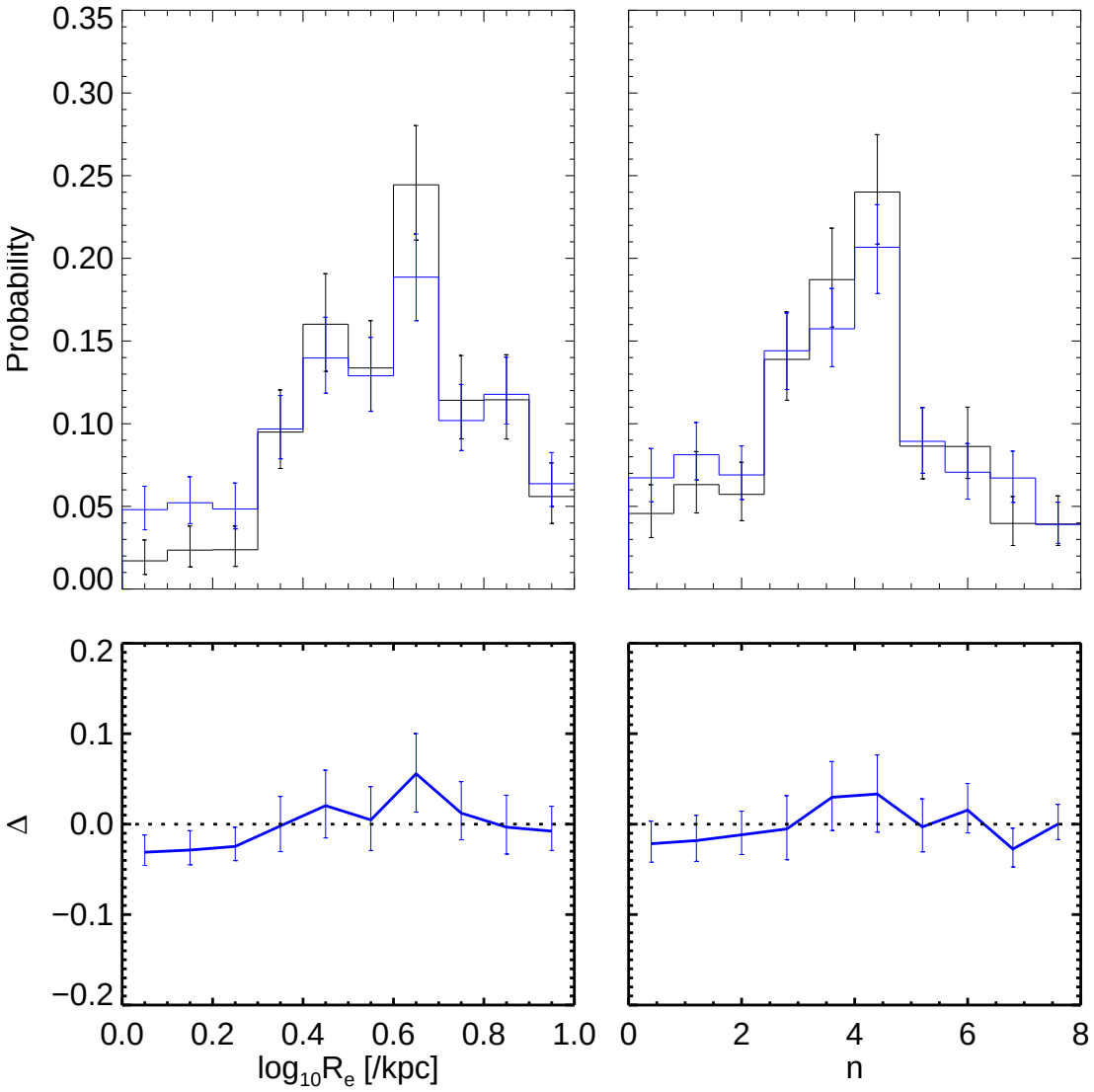


Figure 4.9: Similar to Fig. 4.8 but now showing the results of the 1DHBM (black histogram) and 2DHBM (blue histogram) for the effective radius (left) and the Sérsic index (right) for our one-component galaxies.

Table 4.2: Statistics for the distributions derived from the 1DHBM for each of the population parameters.

Single component galaxies			
		Percentile	
Parameter	**50%**	**16%**	**84%**
μ_e [i-mag arcsec^{-2}]	19.56	18.82	20.52
R_e [kpc]	4.27	2.61	7.23
n	4.00	2.61	7.22
two-component galaxies			
		Percentile	
Parameter	**50%**	**16%**	**84%**
μ_e [i-mag arcsec^{-2}]	18.26	16.54	19.86
R_e [kpc]	1.19	0.55	3.52
n	3.39	1.30	6.06
μ_0 [i-mag arcsec^{-2}]	18.27	17.52	19.73
h [kpc]	3.23	1.88	7.28

tended and fainter than the two-component galaxies (see Table 4.2). We can see in Table 4.2 that the Sérsic parameter distributions differ slightly with the two-component galaxies having typically smaller values. This is clear in the 1DHBM for n seen in Figure 4.10 where there appears to be a bimodality between low and high n values. The distribution of n for the single components also have a small group of low n values, but as stated previously we expect to find some pure discs (i.e., $n \sim 1$) within our one component sample.

The Kormendy Relation (Kormendy 1977; KR hereafter) is a linear scaling relation between the logarithm of the R_e (in kpc) and μ_e (mag arcsec$^-$2):

$$< \mu_e >= \alpha + \beta \log(R_e) \tag{4.17}$$

where we consider the mean surface brightness $< \mu_e >$ measured inside R_e. According to the KR larger galaxies are lower average effective surface brightnesses than smaller ones. The KR has been used to study the structural properties of both ellipticals and spiral bulges (Kormendy, 1977; Bernardi et al., 2003; La Barbera et al., 2003; Gadotti, 2009; Fisher & Drory, 2010; Kim et al., 2016; Costantin et al., 2017). Elliptical galaxies are known to obey a tight linear relation. Kim et al. (2016) also showed that bulges in galaxies with higher B/T values ($0.7 < B/T < 1.0$) closely overlap with the relation derived for ellipticals. However, there are suggestions that there is a departure from the relation for bulges in galaxies with smaller B/T ratios, such that they have a lower mean surface brightness for a given effective radius. Bulges with Sérsic indices close to one are considered to be a product of secular evolution of a bar or disc structure (Carollo 1999; Kormendy

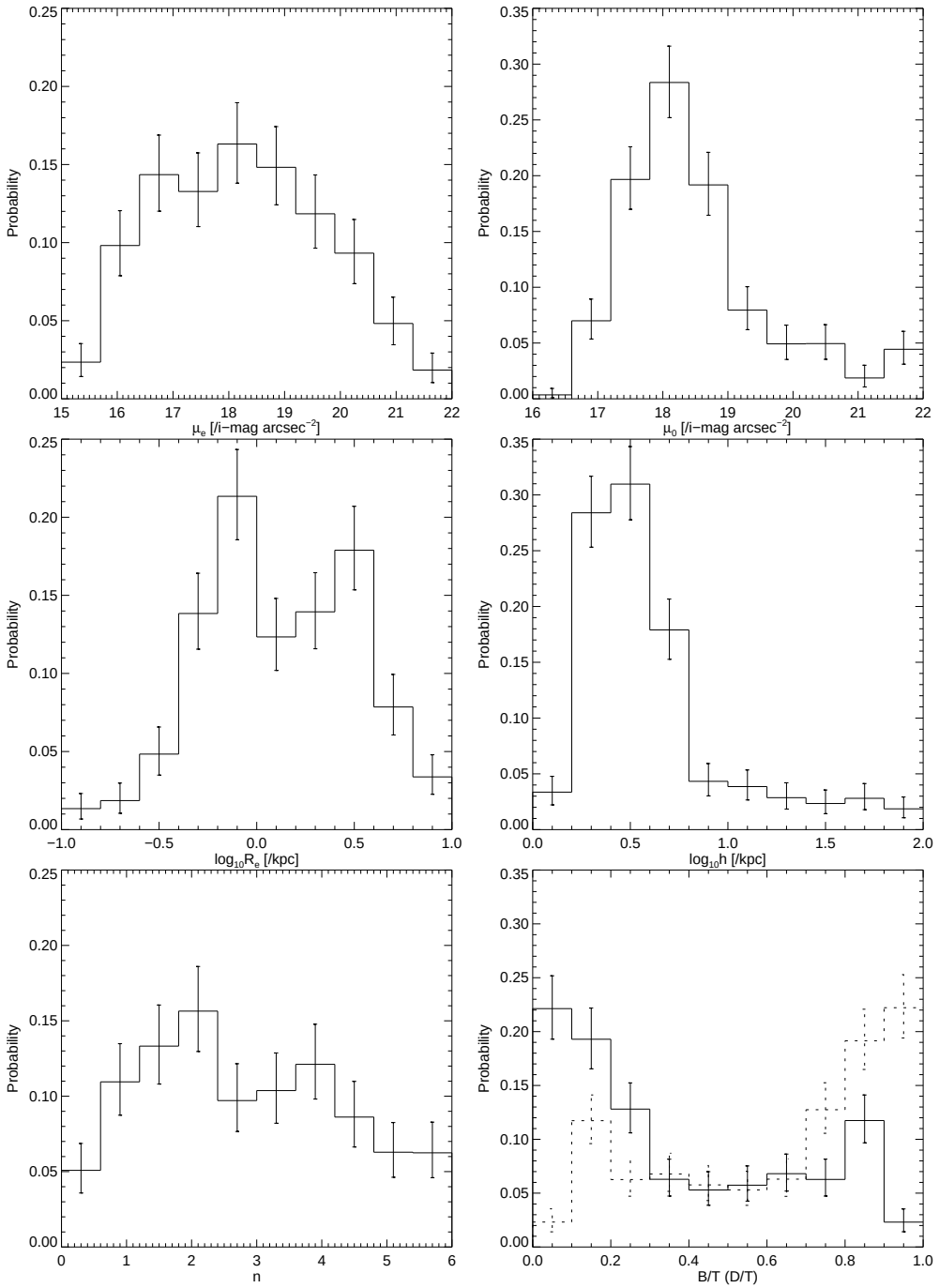


Figure 4.10: Results of the 1DHBM for μ_e, R_e, n, μ_0, h and the bulge-to-total B/T (and disc-to-total D/T shown with the dashed line) ratio for the two-component galaxies in the sample.

& Kennicutt 2004). In Kim et al. (2016) bulges in galaxies with lower B/T values ($0.1 < B/T < 0.3$) seemed more consistent with discy bulges while bulges in galaxies with high B/T seemed likely to be classical bulges. The interpretation has commonly been that bulges have different formation mechanisms depending on their relative size (Silk & Bouwens 1999; Carollo 1999; Kormendy & Kennicutt 2004; Fisher & Drory 2010; Fernández et al. 2014). However, recent studies suggest that the KR is dependent on the mass of the galaxy with the result being that the KR is not a good way to understand bulge mechanisms (Costantin et al. 2017).

Figures 4.11 and 4.12 show the results of the 2DHBM for all the Sérsic parameters of the one and two-component galaxies with the second row showing the KR. Both the one and two-component galaxies in the KR relation follow a similar KR with similar scatter. Even though the single component systems include pure discs there appears to be no significant population that lies off the KR. Similarly there does not appear to be a population lying off the relation, that previous studies have concluded are disc-like bulges (Gadotti 2009).

There are obvious trends in all the panels within Figures 4.11 and 4.12. The subgroup lying off the R_e-n and μ_e-n relations (in the bottom-right corner) for the single component galaxies is likely to be pure disc galaxies (with $n \sim 1$). This can be confirmed with the use of the machine learning classifications as these objects would have a lower probability of being ellipticals. For the rest of the panels there is a smooth transition from smaller, lower surface brightness objects with lower n values to more extended objects with higher n values. Gadotti (2009) suggested that for elliptical galaxies n increases with R_e for bulges but stays reasonably constant for ellipticals, we find instead that there is a positive correlation in the n-R_e plane for both the single and two-component galaxies.

Figure 4.13 shows the relations between the Sérsic profile parameters of the two-component galaxies and the bulge-to-total luminosity ratio. Systems with a low bulge presence over the disc are more likely to be brighter, smaller and have values of n closer to that of a disc. There is no suggestion of multiple populations being present, a smooth transition from low to high B/T. Similar trends have been seen in Gadotti (2009) and Fisher & Drory (2008) but with the interpretation that pseudo and classical bulges occupy different loci of the μ_e, R_e, n-B/T relations. Normally these trends are depicted as being linear but our results suggest differing relations between μ_e, R_e, and n-B/T.

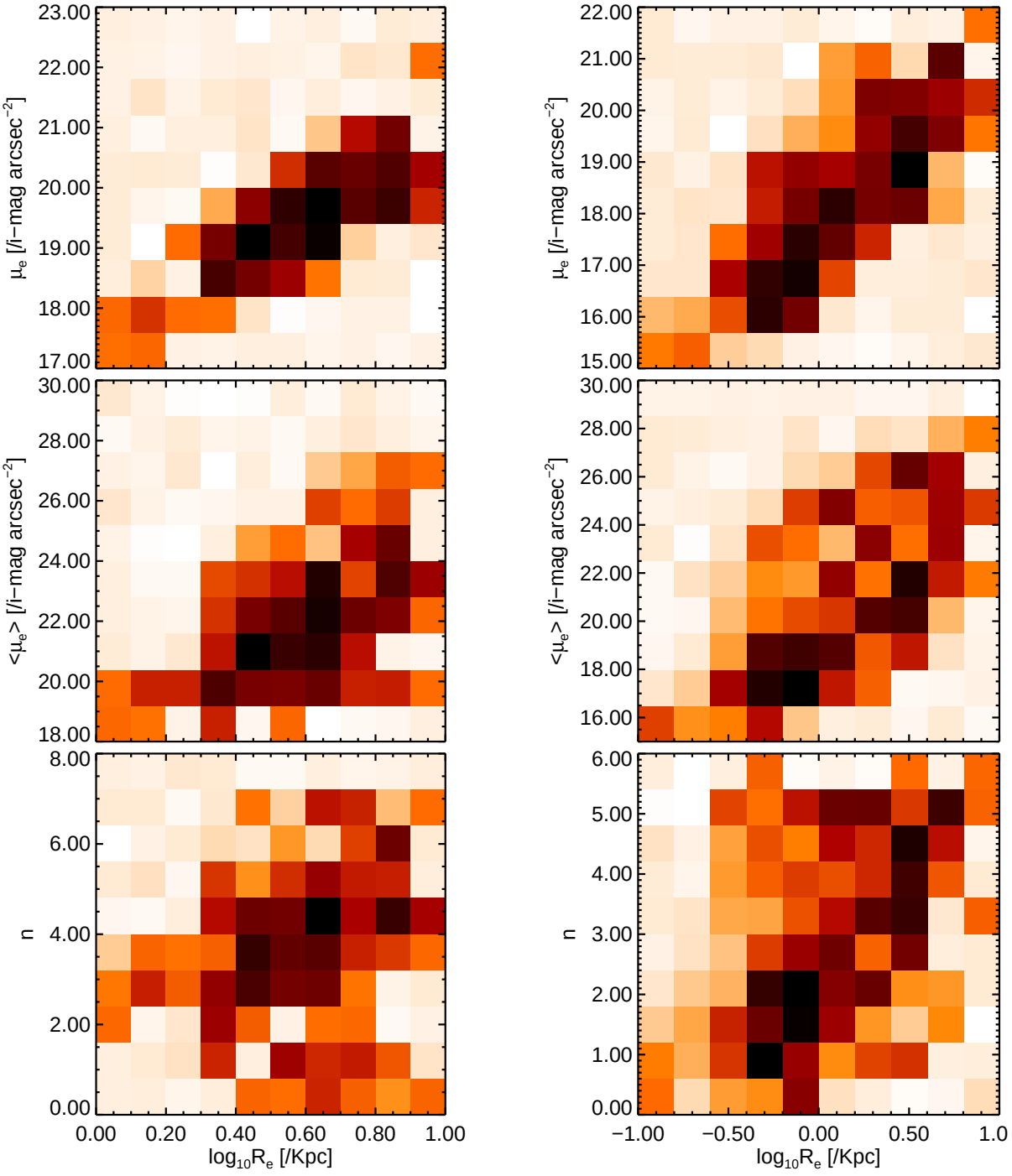


Figure 4.11: The results from the 2DHBM for the Sérsic parameters for the one component galaxies (left) and the two-component galaxies (right). The first row shows μ_e vs. R_e, the second row shows the KR or the mean effective surface brightness vs. R_e and the last row shows n verses R_e. The colours for each bin represent the median posterior probability for the fraction of objects in that bin. A full posterior distribution is calculated for each bin, but not shown here. Refer to the colour bar in Fig. 4.5 for the bin values.

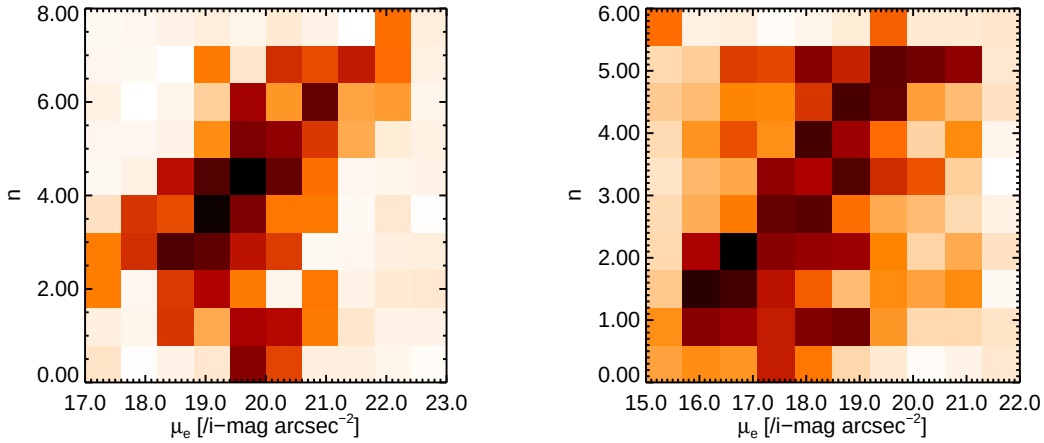


Figure 4.12: Same as Fig. 4.11 but showing μ_e vs. n.

4.4.4 Population distribution of disc structural parameters

Figure 4.14 shows the results of the 2DHBM of the exponential parameters (μ_0 and h). There is a strong correlation between the central surface brightness and the scale length. We can also look back to the 1DHBM results for the same parameters in Figure 4.10. Gadotti (2009) reported on a double peaked distribution in B/T, and consequently in the disc-to-total (D/T) luminosity ratio (D/T maxima at about 0.5 and 0.9). We also find a double peak, however the peak associated with highly bulge dominated galaxies is not as strong as in Gadotti (2009) and our peaks for D/T are at 0.15 and 0.9. The scale-lengths of this work cover a wider range than presented in Gadotti (2009) for the same sample. The median value for h in Gadotti (2009) is 2.8 ± 1.2, which is consistent with the median h found here (3.23, see Table 4.2). However, there is a significant tail of very large, faint discs in our results that is not seen in Gadotti (2009). It is possible that these fits were considered non-physical and discarded or re-fit to find a new minimum. We trust that out MCMC algorithm has properly explored the parameter space and, since all the objects in our sample have fully converged, a cut based purely on our pre-conceptions is not justifiable. Some studies devoted to studying low surface brightness galaxies (for details see Bothun et al. 1997) where there is evidence that galaxies can have large disc scale lengths extending out to as far as 70kpc. Due to the low surface-brightness of these galaxies they are poorly resolved and numbers in this range are very low. Thus these very large low surface brightness discs found in our sample, although very rare, should not be discarded as a fitting error.

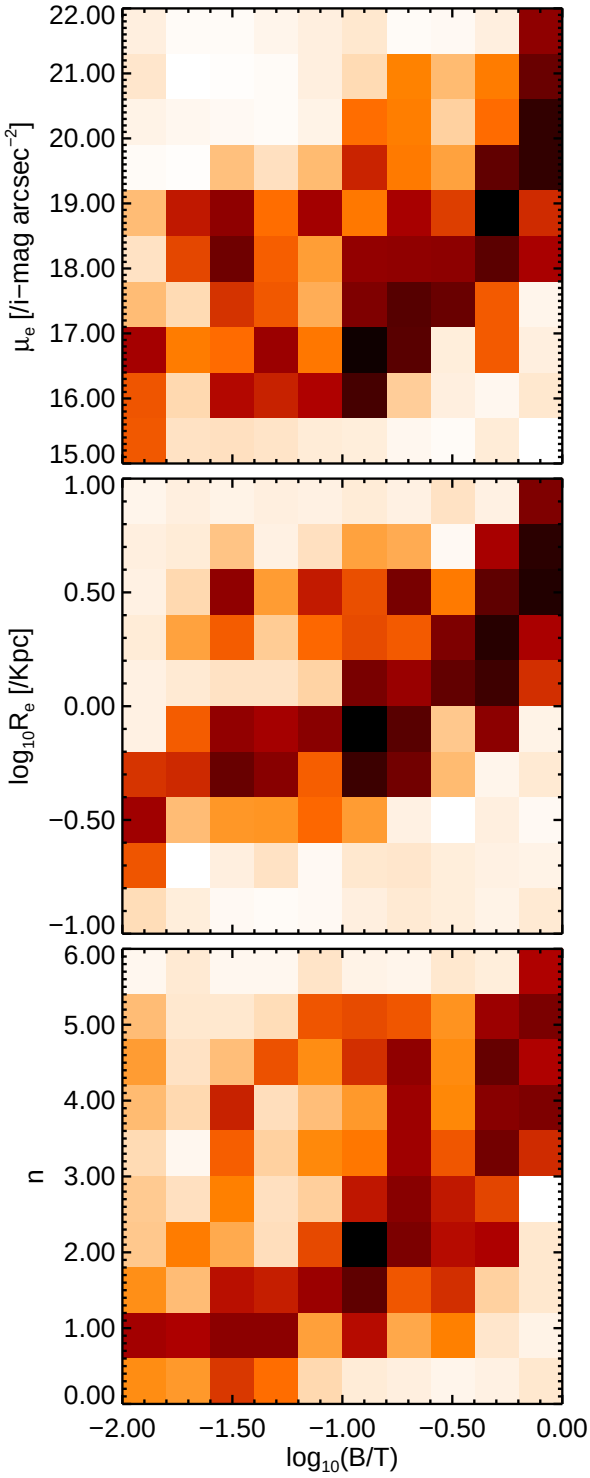


Figure 4.13: Similar to Fig.4.11 showing the bulge-to-total ratio for two-component galaxies plotted against the effective surface brightness,the effective radius and the Sérsic index. They all follow similar correlations with increasing $\log(B/T)$.

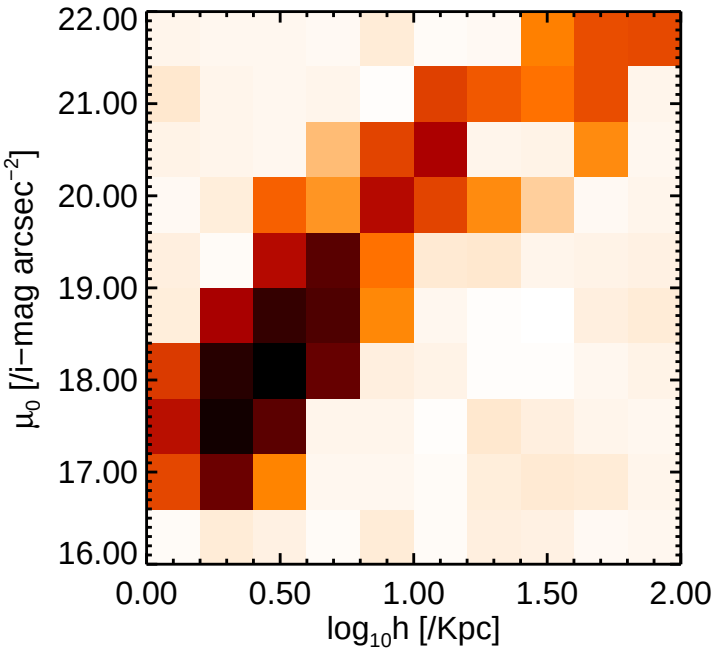


Figure 4.14: The results from the 2DHBM for the exponential profile parameters μ_0 and h.

4.4.5 The population distribution of the bulge-disc interplay

Fig.4.15 shows that the scale-length of the exponential component strongly correlates with the central surface brightness of the disc but correlates poorly with the Sérsic parameters. There is however a subtle correlation with the bulge effective radius suggesting that larger bulges reside in larger discs. This relation has been reported in previous works, but with evidence of a stronger correlation (Couteau et al. 1996; MacArthur et al. 2003; Méndez-Abreu et al. 2008). The priors on the structural parameters (see Sections 2.3.1) do not appear to have an effect on the R_e/h relations as the probability of a galaxy having $R_e/h \sim 1.678$ is low (Fig.4.15). However, looking at Fig. 4.13 we can see that there is a higher probability for a galaxy to have a B/T value close to 1 which is the upper boundary for the prior. After further investigation we find that this is due to the choice to look in the logarithm of B/T and in fact the boundary is rarely hit (see the bottom right panel of Fig. 4.10). For a future study, using bins that vary logarithmically could circumvent this problem.

We find stronger relations between the Sérsic parameters and the ratio between the bulge effective radius and the disc scale-length (right column in Fig. 4.15). This shows that for a galaxy where the size of the disc is dominant over the bulge size it is more likely that the bulge will have a higher surface brightness and have lower values of n.

Through a hierarchical clustering scenario, numerical simulations have shown that major mergers can form bulges and discs producing similar relations seen in the $R_e/h-n$ plane (Scannapieco & Tissera 2003; Tissera et al. 2006; Naab & Trujillo 2006). These simulations of galaxies have shown that the formation of different bulge types (i.e., classical and disc-like) is a function of the galaxy's mass and disc fraction or size of the disc in relation to the bulge. Hence, this is also comparable to our $R_e/h-n$ relation as galaxies with lower R_e/h are more likely to have lower n values expected for secularly evolved bulges.

4.5 Bayesian vs. Machine learning classifications schemes

So far we have focused on our galaxy classification method of using the ΔBIC ($= 25$) to classify the galaxies into one and two-component systems, i.e., it classifies based on the most likely model. As we have seen this has led to some discrepancies when compared to previous studies regarding structures of galaxies in the local Universe. We can, however, classify our galaxies using a different classification proxy to determine whether an object is more likely to be an elliptical or a spiral galaxy. As previously said we have used the results of from Huertas-Company et al (2011) to cut our entire sample on how likely they are to be elliptical (a probability of $P(elliptical) > 0.5$). Separating on whether a galaxy is a elliptical (or not) is not fundamentally the same as separating a sample into galaxies with different number of components. In the ΔBIC method, purely disc galaxies would be considered as single components along with elliptical galaxies.

Figure 4.16 shows the results of the 2DHBM for the Sérsic parameters of the elliptical galaxies and the galaxies classified as being discs (with and without bulges) from the machine learning outcomes of Huertas-Company et al (2011). We can see our hypothesis was correct in the μ_e-n plots for the elliptical galaxies as all the pure discs have now been removed leaving a linear relation between the two parameters. The distributions from the 1DHBM also show this, with no tail to lower values of n for the ellipticals.

4.6 Conclusions

In this chapter, we have analysed the structural components of a sample of 655 SDSS galaxies to investigate their possible scaling relations. The galaxy sample was photometrically decomposed using both one and two-component model as described in Chapter 3. Using this information, together with a HBM, we constrain the mechanisms of bulge and disc assembly in the nearby Universe.

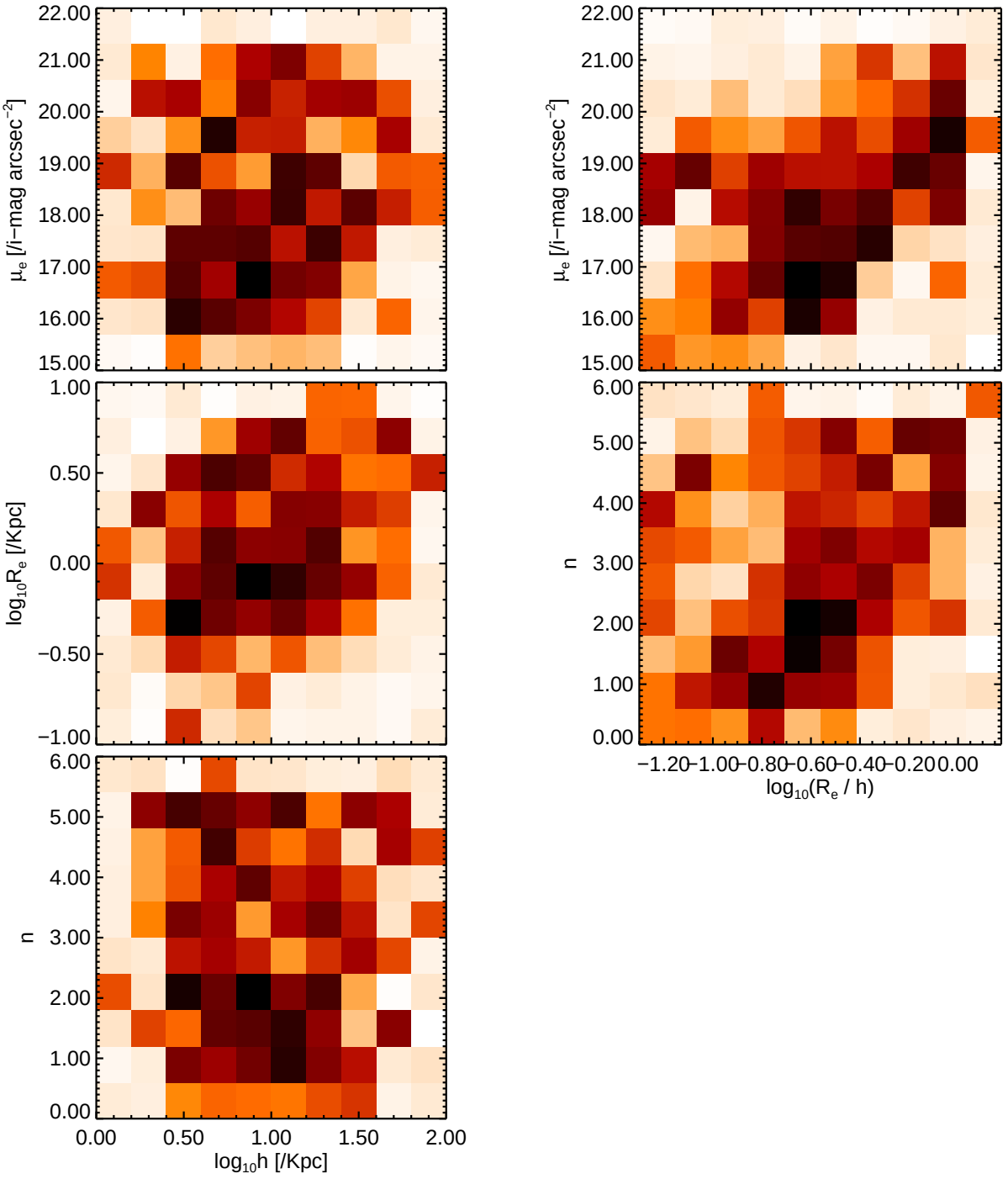


Figure 4.15: The left panel shows the results for the 2DHBM for the Sérsic parameters μ_e, R_e and n verses the disc sclae length h. The right panel shows again the Sérsic parameters μ_e and n but now verse the ratio R_e/h.

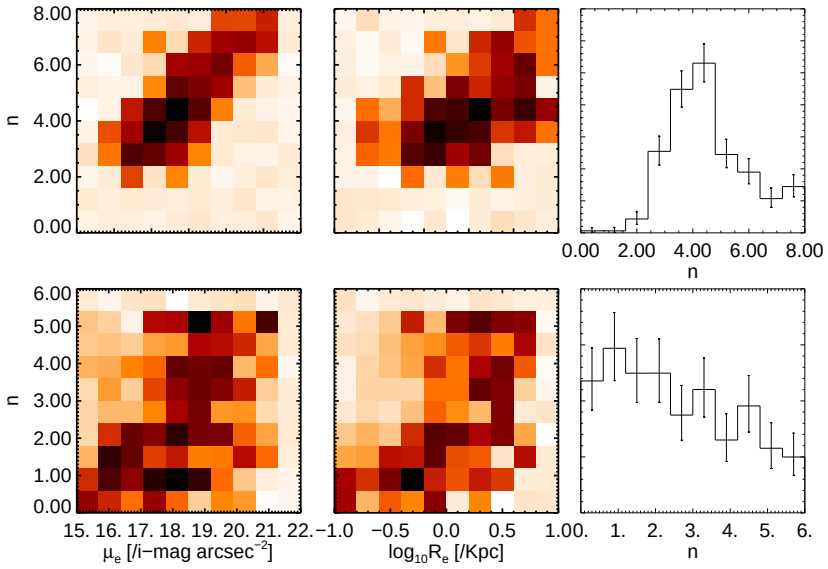


Figure 4.16: The results of the 2DHBM for the Sérsic parameters of the elliptical galaxies (top row) and the disc galaxies (bottom row). The 1DHBM results for n are presented in the histograms with the error bars representing the 16% and the 84% percentiles as before.

- We show how Hierarchical Bayesian models can be used to help estimate global structural properties of the sample. The iterative methodology of population modelling allows us to reduce the uncertainties due to the so-called Bayesian *shrinkage*.

- Using a piece-wise constant representation for the galaxy population we describe the global trends utilising the 2D photometric decompositions of the sample of galaxies from the MCMC algorithm presented in the previous chapters.

- We find that the Sérsic parameters of both the single and two-component galaxies follow similar trends despite their location on the KR or their estimated n value. This suggests that fundamentally the formation of bulges is directly linked to their relative size.

- The only correlation between the Sérsic and exponential parameters that was found in our two-component galaxies was between R_e and h, with large bulges more likely to reside in large discs. However, stronger correlations are present between R_e/h vs. μ_e and n. Therefore, more disc dominated galaxies have a greater probability of having higher surface brightnesses and lower n values closer to that of the disc.

- Our measurements for the global structures of galaxies are consistent with theoretical predictions. However, our results can not distinguish scenarios of a hierarchical growth (*i.e.,* major

mergers) vs. secular evolution. An investigation of these relations at higher redshifts might help to differentiate between the different formation scenarios.

- We have also looked into how the classification of the sample into one- and two-component galaxies influence the results. We compare use of the ΔBIC ($= 25$) threshold to results from machine learning separating the galaxies into pure ellipticals and disc (with and without a bulge) galaxies. We find that when we separate them according to the ΔBIC proxy we include pure disc galaxies in the single components. These are removed when using the machine-learning approach due to the inclusion of other galaxy properties such as the colour.

5

The Structural Evolution of Galaxies since $z \sim 1$ in CANDELS/3D-HST

5.1 Introduction

Although detailed morphological studies of galaxies have been conducted since the early twentieth century, the underlying physical processes responsible for the formation and subsequent evolution of galaxy morphologies remains to be understood. Even with the complex anatomy a galaxy may have, they have mainly been revealed to stick to two broad classification: bulge-dominated or disc-dominated systems. According to the general picture of galaxy formation and evolution in the context of the Λ Cold Dark Matter (ΛCDM) hierarchical scenario, discs form generically inside the evolving CDM haloes, while bulges are mainly grown from the merging of galaxies or due to intrinsic disc instabilities (see the reviews by Brooks & Christensen 2016 and Bournaud 2016 for more details).

Studies have also shown a strong colour-morphology bimodality containing populations of red, bulge-dominated (*e.g.* elliptical) and blue, disc-dominated (*e.g.* spirals) galaxies (Baldry et al. 2004;

Driver et al. 2009; Drory & Fisher 2007). This colour classification correlates strongly with the star formation (SF) of a system (Baldry et al. 2004; Brinchmann et al. 2004; Kauffmann et al. 2003; Ilbert et al. 2013). The red, bulge-dominated galaxies have insignificant or very little SF and the blue, disc-dominated galaxies are said to lie on the star formation main sequence (we refer to it here as the star formation rate - stellar mass relationship, SFR$-M_*$); a positive correlation between SF and the stellar mass (Noeske et al. 2007; Daddi et al. 2007; Wuyts et al. 2011; Rodighiero et al. 2011). Recently however there have been an increasing number of findings presenting blue bulges and red discs (Bamford et al. 2009; Master et al. 2010). These objects put into question the merger-driven hierarchical evolution paradigm altogether. The formation scenarios that lead to these structures (and the properties they exhibit) we see in the local Universe is, as always, as broad and complex as the galaxies themselves.

Increasingly detailed morphological studies are now able to push further back in cosmic time into the epoch where the star-formation rate density of the Universe peaks (at around $1 < z < 3$). These studies have suggested that the Hubble sequence is present at these high redshifts, however a larger fraction of disc-dominated systems (Buitrago et al. 2013; van der Wel. 2011; Chevance et al. 2012; McLure et al. 2013; Bruce et al. 2014) as well as systems that are morphologically disturbed are in abundance (Mortlock et al. 2013). Also present is the SFR $- M_*$ relation which has been reported to extend down to redshifts around $z \sim 2.5$ (Wuyts et al. 2011). The SFR $- M_*$ relation's connection with the morphology also holds until this time with starforming objects often having lower Sérsic indexes (i.e. discy morphologies). These works however have largely relied on single component fits to the galaxies (usually with a Sérsic profile), which has given an insight into their morphology but can leave out hidden structural relationships, such as how the bulges within disc galaxies can influence the overall evolution.

This chapter is organised as follows: the CANDELS/COSMOS field as well as the 3DHST project is first introduced followed by our description of how we determine the progenitors of galaxies in the near-by Universe. In Sect. 5.5 we give an overview of how we have applied `PHI` as well as the Hierarchical Bayesian Model described in Chapter 4 to obtain structural parameters for the population of local and high redshift galaxies. In Sect. 5.6 we then out line the results describing the structural evolution of one- and two-component galaxies. We finish this chapter with a summary of the results and a discussion on how they compare to previous observational and theoretical studies.

5.2 CANDELS and the COSMOS field

Over the last twenty years our understanding of the formation and the evolution of galaxies has advanced enormously due to the combined efforts of both theoretical and observational efforts. The *Hubble Space Telescope* (HST) along with the largest ground-based telescopes have been used to perform deep observational studies to probe galaxy populations back to around one billion years after the beginning of the Universe.

The addition of the Wide Field Camera 3 (WFC3) / Infrared (IR) instrument on HST fostered the most ambitious near-IR surveys that HST had conducted (Grogin et al. 2011; Koekemoer et al. 2011). The WFC3/IR enables a two-filter survey with an area of $\sim$ 600arcmin^2 to limiting magnitudes of $H_{AB} \gtrsim 27.1$ and $J_{AB} \gtrsim 27.0$ (The H_{AB}/F160W filter has a zeropoint magnitude of 25.96 and the J_{AB}/F125W filter of 26.25). The usable portion of the detector is 1024 × 1024 pixels, covering a region of 136″ × 123″ with a native pixel scale of 0.128″ pixel^{-1}. The observations are carried out in the wide filters: *F125W* and *F160W*, which cover the wavelength ranges of $\sim 1.1\mu m - 1.4\mu m$ and $\sim 1.4\mu m - 1.7\mu m$, respectively. The WFC3/IR Point Spread Function (PSF) has a Full Width Half Maximum (FWHM) between 0.13″ and 0.15″ over this wavelength range. Having the capacity to observe longer wavelengths with a large-area allows HST to follow galaxies well into the reionisation era, and to observe morphologies of galaxies down to $z \sim 2$ when the cosmic star formation density peaks and the Hubble sequence was starting to emerge.

The observing program that utilises on the WFC3/IR is the Cosmic Assembly Near-infrared Extragalactic Legacy Survey (CANDELS) which has five distinct fields; The Great Observatories Origins Deep Survey North and South (GOODS-N and GOODS-S; Giavalisco et al. 2004), the Extended Groth Strip (EGS; Davis et al. 2007), the UKIDSS Ultra-deep Survey (UDS; Lawrence et al. 2007) and COSMOS (Scoville et al. 2007). With each field having two distinct depths. The CANDELS wide portion is used in all five fields and the deep portion is only used in the GOODS-N and GOODS-S.

The three wide fields (i.e., COSMOS, UDS, and EGS) consist of a continuous mosaic of overlapping WFC3/IR titles (Figure 5.1 shows the COSMOS field where the blue regions highlight the WFC3/IR tiles). The exposures are taken over two HST orbits with roughly twice as much time being allocated to the F160W filter.

For our extragalactic studies we have utilised the WFC3/IR images in the COSMOS field. The COSMOS field is a 2-degree region of sky surveyed with HST (Scoville et al. 2007; Koekemoer et al. 2007). The field is centred at α(J2000) = $10^h\ 00^m\ 28^s$ and δ(J2000) = +02° 12 ′ 21″. The

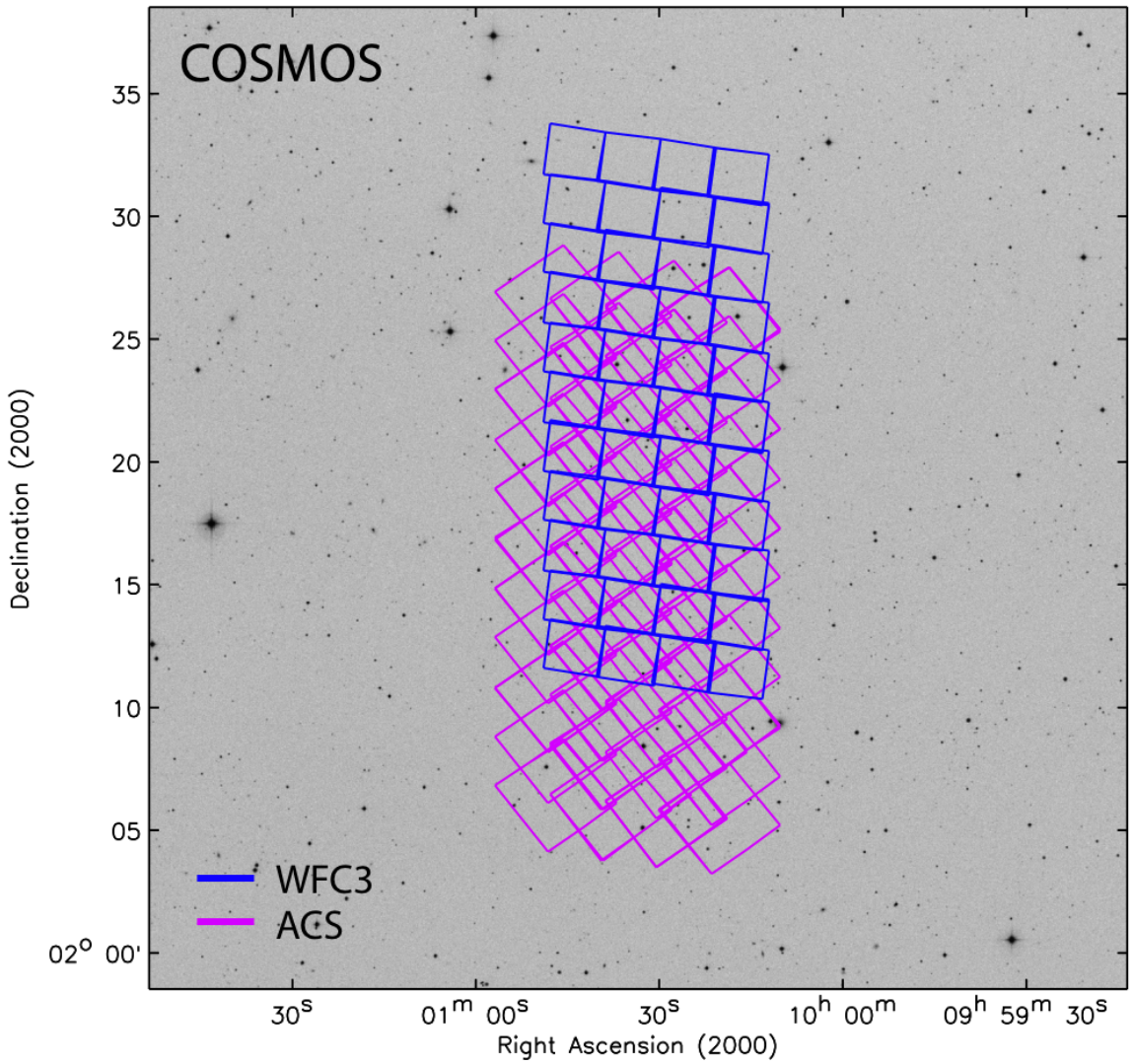


Figure 5.1: Footprint of the CANDELS observations in the COSMOS field with WFC3/IR prime exposures shown in blue and ACS/WFC parallel exposures shown in magenta. (From Grogin et al. 2011)

CANDELS WFC3/IR observations in the COSMOS field consist of a rectangular grid of 4×11 tiles ($\sim 8.6' \times 23.8'$) that are placed north to south with a a 180° position angle. The layout of the mosaic within the COSMOS field is shown in Figure 5.1.

5.3 The 3D-HST project

Building on the CANDELS survey, the WFC3 spectroscopic survey in the same fields is the 3D-HST project (Skelton et al. 2014). 3D-HST is a 248-orbit HST treasury program that uses the WFC3 grism for slitless spectroscopy that covers roughly three quarters of the CANDELS area. The 3D-HST project has provided the community with a wealth of information in the CANDELS/3D-HST fields. Using SExtractor (Bertin & Arnouts 1996) objects have been categorised from deep combined WFC3/IR images. The 3D-HST project provided newly reduced images, catalogues, PSFs ,*etc.* for all the the CANDELS fields. These source catalogues along with detection images and associated segmentation maps and PSFs, have been used to measure publicly available imaging data sets. The images are given with the background subtracted but we have checked for zero sky by using the segmentation maps and removing all the identified objects, we then calculate the mean and standard deviation of the background noise. This has resulted in high quality Spectral Energy Distributions (SEDs) where photometric redshifts have been obtained. Stellar populations parameters have then been obtained from fitting the SEDs with stellar population synthesis models using the photometric redshifts as inputs. For our work we use the imaging data provided by the 3D-HST project as well as their photometric catalogues containing both the redshift and stellar population parameters.

5.3.1 Redshifts and stellar mass estimates

Photometric redshifts, stellar masses, and star formation rates (SFRs) are estimated from the SED modelling as described by Wuyts et al. (2011, 2012) and Barro et al. (2013, 2014). The redshifts used for the COSMOS sample in this chapter are taken from the catalogues given by the 3D-HST team (Skelton et al. 2014). We describe here the basic procedure and refer the reader to the reference work for more details. Photometric redshifts are estimated from a variety of different codes available in the literature which are then combined to improve performance. The technique is fully described in Dahlen et al. (2013). Photometric redshifts were determined from fitting the SED of each object with combinations of galaxy templates using the EASY code (Brammer et al. 2008). The default template set, described in Brammer et al. (2008), was used which includes the five templates derived from a library of PEGASE stellar populations synthesis models (Fioc & Rocca-Volmerange 1997), a young, dusty template and an old, red galaxy template as described in Whitaker

et al. (2011). Based on the best available redshifts (spectroscopic or photometric) stellar masses and SFRs are then estimated using FAST (Kriek et al. 2009) assuming Bruzual & Charlot (2003) models, a Chabrier (2003) IMF, solar metallicity, exponentially declining star formation histories, and a Calzetti et al. (2000) extinction law.

5.4 Sample selection

For our local galaxy population we have used the same SDSS galaxy images predented in earlier chapters (see Section 3.4). We have used the HST WFC3/IR data from the CANDELS and 3D-HST multi-cycle treasury programme (Grogin et al. 2011; Koekemoer et al. 2011) centred on the COSMOS field (Scoville et al. 2007; Skelton et al. 2014). According to the position of the centroids for each galaxy provided by the 3D-HST catalogues, we take cut-outs from the *J* (F125W) and *H* (F160W) filter images. Corresponding weight maps and segmentation maps are also cut using the same centroids and sizes. For the SDSS sample, bar galaxies (previously identified in Gadotti (2008)) were removed from the final sample. For the COSMOS sample we have visually removed galaxies with obvious bars but as the bar fraction decreases with redshift (Sheth et al., 2008) the problem of finding a bared galaxy is rare.

5.4.1 Redshift range and morphological *k*-correction

Galaxies have very different structures depending on what wavelength range they are observed at (Bohlin et al. 1991; Kuchinski et al. 2001; Windhorst et al. 2002; Papovich et al. 2003; Conselice 2004). This shift in the qualitative and quantitative appearance of a galaxy constitutes as a morphological k-correction. A primary problem in understanding the evolution of galaxy structures over cosmic time is constraining the effects of the morphological k-correction.

Most deep high resolution imaging is done in the observed-frame optical wavelength range, probing up to $\lambda \sim 1\mu m$, and allowing for a sampling of the rest-frame optical (> 4000Å) wavelength range only up to a redshift of $z \sim 1.5$ (Conselice 2004). Above this redshift, the rest-frame ultraviolet begins to be observed where predominantly young stars with ages < 100 Myrs are sampled. With that in mind, the largest differences would be observed in galaxies composed of young and old stars which are not spatially mixed, such as early type spiral galaxies (see Windhorst et al. 2002).

Morphological k-corrections depend on both the galaxy type and redshift. In Huertas-Company et al. (2009) the effects of this morphological k-correction at $1 < z < 2$ was quantified by comparing morphologies in the K and I-bands in the COSMOS field. Their classification was based on

a machine learning approach quantifying the galaxy morphology on non-parametric values. They found that I-band classifications are less able to identify early-type galaxies than the K-band, thus underestimating the elliptical population.

With the above information in mind, it is clear that selecting the appropriate redshift range to probe the same rest-frame wavelength range within different filters at higher redshifts is fundamental to avoid any misconceptions in interpreting our results due to morphology k-correction. Thus, to properly compare our local sample ($0.04 < z < 0.06$) in the SDSS i-band to the two HST WFC3/IR band filters we use images in the redshift ranges $0.5 < z < 0.85$ and $0.85 < z < 1.27$ for the J and H bands respectively. The redshift ranges are defined when the effective wavelength of the i-band is traced past the minimum and maximum wavelengths of both the J and H band.

5.4.2 Selecting the progenitors of local galaxies

Evolutionary processes, make linking a galaxy in the local Universe with its progenitors difficult. When inferring the evolution of individual galaxies and linking them with progenitors without being strongly affected by progenitor bias is a key issue many previous studies have devoted time in solving (Carollo et al. 2013; Sonnenfeld et al. 2014; Shankar et al. 2015). It has been shown that the stellar mass function (SMF) has evolved significantly since $z \sim 3-4$ (Pérez-González et al. 2008; Muzzin et al. 2013; Ilbert et al. 2013) so a sample selection based on a fixed stellar mass can be affected by galaxies entering in at lower redshifts. Another method to link progenitors or descendants is by linking galaxy populations at a constant luminosity (Wake et al. 2006) or by isolating specific galaxy populations such as bright cluster galaxies (Lidman et al. 2012; Lin et al. 2013; Shankar et al. 2015). A popularised progenitor linking method in recent years is the selection using a fixed number density (van Dokkum et al. 2010; Bezanson et al. 2011; Conselice et al. 2013; Patel et al. 2013; Mundy et al. 2015; Torrey et al. 2016). The underlying assumption is for example, massive galaxies at higher redshifts will evolve into massive galaxies in the local Universe. As galaxy masses and other properties are thought to evolve significantly over time, the number density is thought to stay reasonably static *as long as the mass rank order among a galaxy population is preserved* (Torrey et al. 2016).

We wanted to confirm the likely mass evolution of a Milky-Way like galaxy (as well as a $M_* = 10^{10} M_\odot$ galaxy) with a toy model consisting of a bulge and a disc component. Using a set of stellar population models we randomly traced a galaxy of mass $\log M_*/M_\odot = 10.66$ back to a redshift of $z = 2$. We used Star Formation Histories (SFHs) from stellar population synthesis models (Bruzual

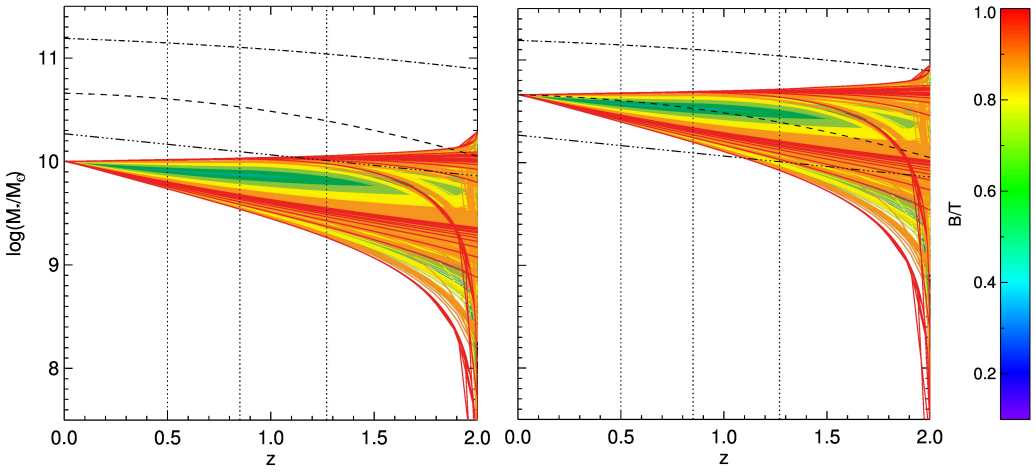


Figure 5.2: The mass evolution of galaxy populations is shown tracked from a redshift $z = 2$ to $z = 0$. The coloured bands are mass estimates from stellar population models combined according to a bulge-to-total ratio so that the result is matches a galaxy with $\log M_*/M_\odot = 10$ (left panel) and Milky-Way sized ($\log M_*/M_\odot = 10.66$; right panel) galaxy at $z \sim 0$. The stellar populations are randomised for each component of the galaxy from: a single stellar burst model, an exponentially declining SFH and a constant SFH. The dashed line is the mass evolution function for a $\log M_*/M_\odot = 10.66$ Milky-Way type galaxy from van Dokkum et al. (2013). The dashed-dotted lines are the mass evolution functions for galaxies with mass $\log M_*/M_\odot = 10.27, 11.2$ from Ilbert et al (2013) and Patel et al (2013) respectively. The vertical dotted lines indicate the redshift ranges obtained from the morphological k-correction analysis described in the above text.

& Charlot 2003) of a single burst, and exponentially declining and a constant SFH. We also altered the bulge-to-total mass ratio for each toy model. Figure 5.2 shows the stellar mass evolution of a $\log M_*/M_\odot = 10$ and $\log M_*/M_\odot = 10.66$ galaxy out to $z = 2$ comparing it to the mass evolution function from a fixed number density approach from van Dokkum et al. (2010). We also varied the formation time of the systems between the redshifts $z \in \{2 : 8\}$. Albeit with a significant scatter at $z \gtrsim 1$, the mass-redshift function of van Dokkum et al. (2010) matches our toy model. We want to stress here that these toy models encompass our ignorance in how nearby galaxies can be traced back through time.

However, there are two problems with the comoving number density analysis (and subsequently our toy models): 1) galaxy mergers will change the total number density of galaxies (Ownsworth et al. 2014) and 2) stochastic growth rates, this makes a galaxy's evolution more of a random process. Torrey et al. (2016) studied the number density evolution of galaxy populations was tracked through the cosmological hydrodynamical simulation Illustris (Vogelsberger et al. 2014; Genel et al. 2014; Nelson et al. 2015). This tracked number density evolution incorporated both the impact of scattered growth rates and galaxy mergers. In a following paper, Wellons & Torrey (2017) then capture how properties of galaxies within the simulations evolved over time and showed that the constant number density method performs poorly in recovering the evolution of galaxy properties. Wellons & Torrey (2017) suggest that their probabilistic number density method works best for tracing galaxy progenitors. Comparing Figure 5.2 to the mass evolution presented in Wellons & Torrey (2017), we find that we over estimate the evolution over time, as expected, due to the random selection of the SFHs for the galaxies.

5.4.3 Final sample

Figure 5.3 shows the final redshift and stellar mass bins for our COSMOS sample. We use the toy models as a rough indicator of how galaxies might evolve. Our local Universe sample in SDSS has a minimum mass limit of $\log M_*/M_\odot = 10$, and a peak in the mass distribution $\log M_*/M_\odot \sim 10.6$, which is conveniently close to the mass of the Milky-way. This motivated us to split the local sample in two mass bins, so that we have two populations of galaxies, with either less than the mass of the Milky-Way or greater than. We then use the toy models to determine a minimum mass in both redshift ranges defined by evolving a galaxy with $M_* = M_{\rm MW}$ and a galaxy with $M_* = 10^{10} M_\odot$ (minimum mass in SDSS sample) back to define their likely progenitor masses (see Fig. 5.2). The blue boxes in Fig. 5.3 represent the samples that will always evolve to masses less than the mass of the Milky-Way and vice verse for the red boxes. The green boxes in Fig. 5.3 are then the sample of

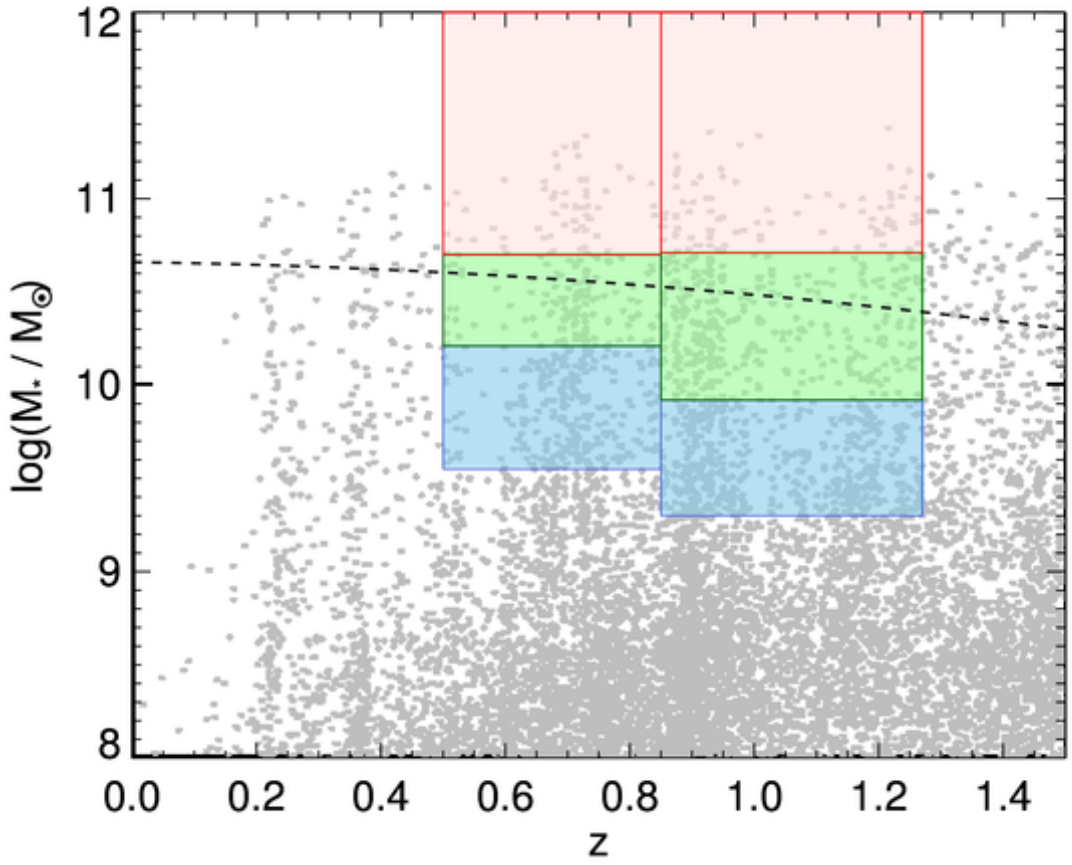


Figure 5.3: Stellar mass as a function of photometric redshift for the galaxies in the COSMOS field described in the catalogues of the 3D-HST project. The boxes indicate our cuts in mass and redshift according to the morphological k-correction of the SDSS i-band filter to the CANDELS WFC3/IR filters and the mass evolutions from the toy models. The blue boxes show galaxy masses that will evolve to become galaxies with masses $\log M_*/M_\odot < 10.66$ at $z \sim 0$, and vice versa for the red boxes i.e. $\log M_*/M_\odot > 10.66$. The green boxes show galaxies that have the potential to evolve to become $\log M_*/M_\odot = 10.66$ or Milky-Way like masses. Mass ranges are all estimated from the toy models explained in the above text. For reference, the dashed line is the mass evolution function for a $\log M_*/M_\odot = 10.66$ Milky-Way type galaxy from van Dokkum et al. (2013).

Table 5.1: Table containing the number of galaxies in each redshift and mass bin

$0.02 < z < 0.07$	N_{gal}
$10 < \log(M_*/M_\odot) \leq 10.66$	232
$\log(M_*/M_\odot) \geq 10.66$	222
$0.5 < z < 0.85$	
$9.55 < \log(M_*/M_\odot) \leq 10.21$	272
$10.21 < \log(M_*/M_\odot) \leq 10.7$	128
$\log(M_*/M_\odot) > 10.7$	60
$0.85 < z < 1.27$	
$9.3 < \log(M_*/M_\odot) \leq 9.92$	487
$9.92 < \log(M_*/M_\odot) \leq 10.71$	336
$\log(M_*/M_\odot) > 10.71$	85

galaxies that can evolve to have a mass $\log M_*/M_\odot = 10.66$. These galaxies are combined within the high-mass sample. The total number of galaxies for our sample in COSMOS is $N_{gal} = 1368$ where the numbers for each redshift and mass bin are shown in Table 5.1 also included is the low-z sample from Chapter 3.

5.5 Bulge-disc photometric decompositions

In this Section, we have used the photometric decomposition code `PHI` described (see Chapter 2 for a full description), which was also used to analyse the low redshift sample in Chapter 3. Thus, the morphologies of the 1368 objects from the COSMOS field, combined with the 454 objects from SDSS, have been fitted with both single and two-component models (either a single Sérsic or a Sérsic + exponential photometric model) making a total of 1793 objects. Figure 5.4 shows three examples of the results from the MCMC fit.

The procedure makes use of the PSF provided in the 3DHST data release v4.1.5. The PSF was created for each HST image by stacking a number of isolated stars across the image. The HST ACS and WFC3 images were convolved to match the F160W image, which has the broadest PSF. Weight maps were also provided and were created by using the empty apertures method. By placing many 0.7" aperture at random points across the image, removing any that overlap with the segmentation map provided after a run through SExtractor (Bertin & Arnouts 1996), and then fitting distribution of resulting aperture fluxes to estimate the error and the weight ($1/\sigma^2$, where σ is the error calculated) in each band. `PHI` is then run on 250×250 pixel cut-out image stamps which were obtained using the centroids given in the 3D-HST catalogues (the size of the image stamp was chosen to ensure that the entire galaxy was cut-out). The initial conditions for each run were estimated from fits to

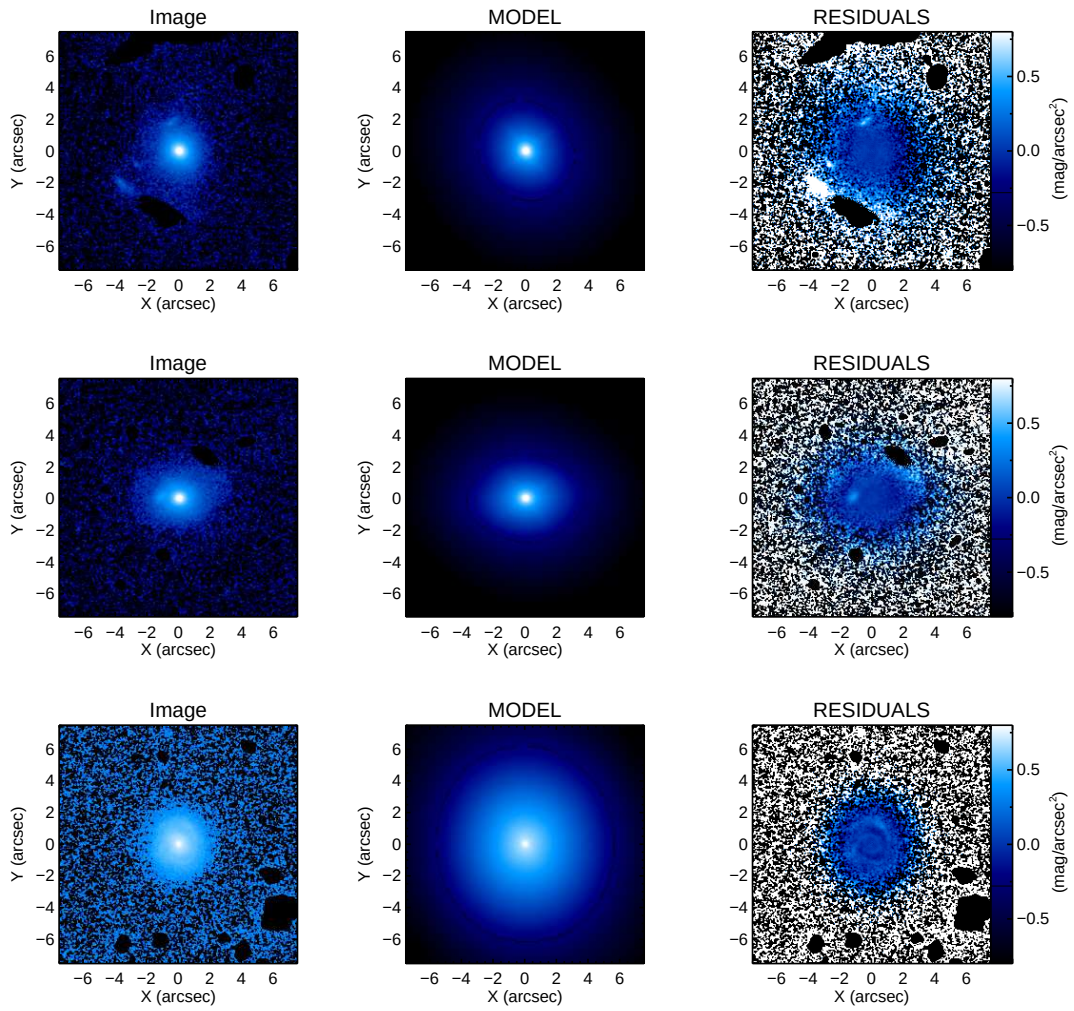


Figure 5.4: Results from the MCMC decompositions for three galaxies from the COSMOS sample. The images are placed in order of their corresponding ΔBIC value. The top row has $\Delta\text{BIC} \leq 0$, the middle row has $\Delta\text{BIC} \sim 0$, and the bottom row has $\Delta\text{BIC} \geq 0$. The top row shows the data (left), the Sérsic only model fit (middle) and the residual. The second row shows the data (left), the bulge-disc model fit and its corresponding residual. The bottom row shows the bulge-disc model fit and its corresponding residual. All the models were made using the medians from the posterior distributions.

the 1D light profile provided after running the ellipse fitting task of IRAF. Each object was fitted with both a single Sérsic and a Sérsic + exponential photometric model. To ensure that the final posterior probability distribution was fully explored we ran three simultaneous chains in the final sampling phase of `PHI` and ended each run when all three chains satisfy the convergence tests implemented in the code (see Chapter 2).

We then used the ΔBIC as a proxy to determine the most probable model for an object. A

ΔBIC > 25 cut for the SDSS sample was used to obtain the most probable two-component systems and thus the same was used for the COSMOS sample. However, it is worth noting that the ΔBIC distribution tightens around ΔBIC $= 0$ as redshift increases and for future studies and full test on simulated data would be needed to distinguish whether this effect is due to galaxies being more single components at high redshift or that the resolution is too low to be able to properly determine a second component.

After conducting our detailed photometric decompositions and determining the most probable model for each object we pass the results through the Hierarchical Bayesian Model (HBM) described in Chapter 4. To understand how the population distribution of the structural parameters have evolved from $z \sim 1$ to $z \sim 0$. To improve the representative of our results, we have taken the Bayesian histogram models and plotted the median probabilities for each bin in increasing probability contours. This gives us the change to directly compare between the different mass groups and redshift ranges.

5.6 Results

Figure 5.5 shows the ratio between the number of one-to-two-component galaxies in the specific redshift and mass bins. The blue open circles represent the lower mass samples (i.e., galaxies that according to our models are likely to evolve becoming a galaxy with $\log(M_*/M_\odot) \leq 10.66$ at $z \sim 0$, the blue boxes in Fig. 5.3) and the red closed circles show the higher mass sample (green and the red boxes in Fig. 5.3). We also compare to the results of Margalef-Bentabol et al. (2016) who did bulge-disc decompositions of a sample of $\log(M_*/M_\odot) > 10.0$ from the CANDELS survey in a redshift range of $1 < z < 3$.

We find that for the low mass sample, the number of single component galaxies dominate at around $z \sim 1$ but then decrease until they match the number of two-component galaxies in the local Universe. A similar behaviour is observed for the high-mass sample, but only when we add the results at higher redshift of Margalef-Bentabol et al. (2016). From our analysis, it appears that single and two-component galaxies were fairly one-for-one at $z \sim 1$ and then again at $z \sim 0$ with a dip at $z \sim 0.7$. The increase in the fraction of two-component galaxies with redshift was also reported in Margalef-Bentabol et al. (2016). Adding the results from Margalef-Bentabol et al. (2016), we can see that the high mass sample could follow the same trend as the low mass sample but at earlier stages in the Universe. This trend can be affected by the signal-to-noise of a galaxy and redshifting effects. Margalef-Bentabol et al. (2016) demonstrated that this was not the case but a full analysis

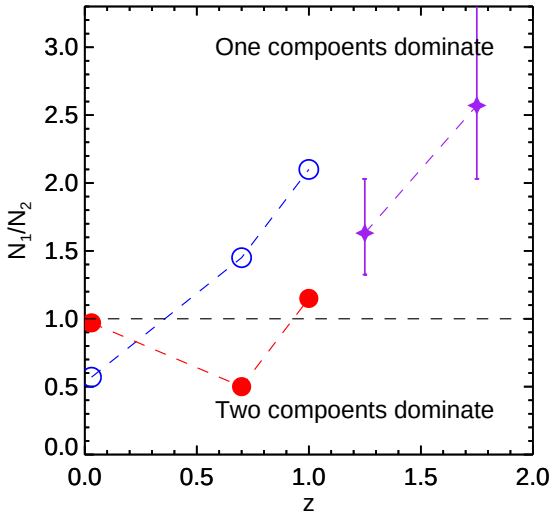


Figure 5.5: Figure showing the ratio between the number of single and two-compoent galaxies for the three redshift ranges (median values for the redshift bins are shown) and two mass bins in the SDSS and COSMOS sample. The blue open circles show the lower mass samples (*i.e*, galaxies according to our toy models are likely to evolve to become $\log(M_*/M_\odot) < 10.66$, blue boxes in Fig. 5.3) and the red closed circles show the higher mass sample (green and the red boxes in Fig. 5.3). The purple stars are the first two data points from Margalef-Bentabol et al. (2016) in their Figure 8 where all galaxies have a mass $\log(M_*/M_\odot) \geq 10.0$.

with synthetic galaxies is needed to properly understand these possible biases.

Figure 5.5 also supports our choice of separating the sample in two mass bins using the stellar population toy models as there appears to be a significant difference between the two groups. We can now study how the overall morphologies of the one and two-component galaxies evolve with redshift. We separate the following sections into one- and two-component galaxies to investigate their individual structural changes over the redshift range.

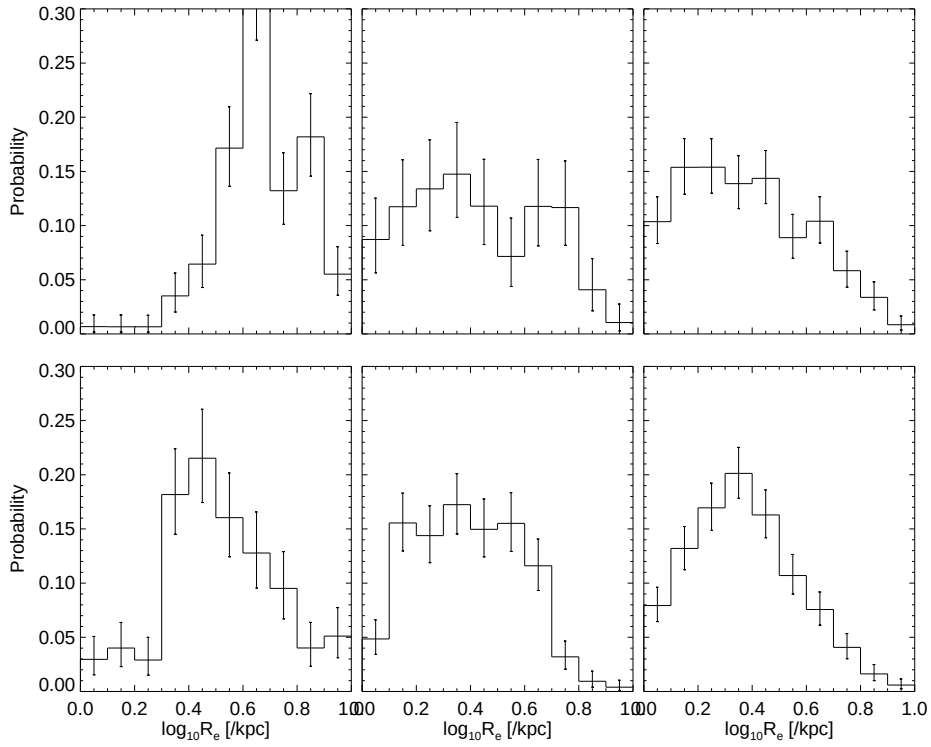


Figure 5.6: The results from the 1DHBM for the R_e of the most probable single component galaxies. The top row shows results for the high mass sub-sample and the bottom row show the lower mass galaxy populations for the three redshift bins $z \sim 0, 0.7$ and 1 (left to right). The values of the histograms are the median values for the fractional counts for each bin estimated with the 1DHBM, and the error bars show the 16th and 84th percentiles.

5.6.1 Single component galaxies

Fig. 5.6 shows the results from the 1DHBM for the R_e population distribution of single component galaxies (ΔBIC < 25). We find subtle differences between the two mass samples over the redshift range. At $z \sim 1$ it appears that the low and high mass sample have similar mean values for R_e but the probability distribution for the high mass systems appears flatter. Both mass groups evolve in size from $z \sim 1$ to $z \sim 0$ with the higher mass probability distribution always peaking at larger radii than the lower mass systems.

For both mass samples, the distribution of n is very similar at $z \sim 1$, as shown in Figure5.7. As we move down in redshift the two mass samples begin to differ. The low mass sample has a gradual increase in n from $z \sim 1$ to $z \sim 0.7$, whereas the high mass sample evolves in n at a much faster pace. At $z \sim 0$ the distribution for the low mass sample resembles the distribution for the high mass sample at $z \sim 0.7$, with a broad peak at high n and a residual probability of smaller n systems around $n = 1$. For the high mass sample there appears to be almost zero probability for systems with $n < 2$.

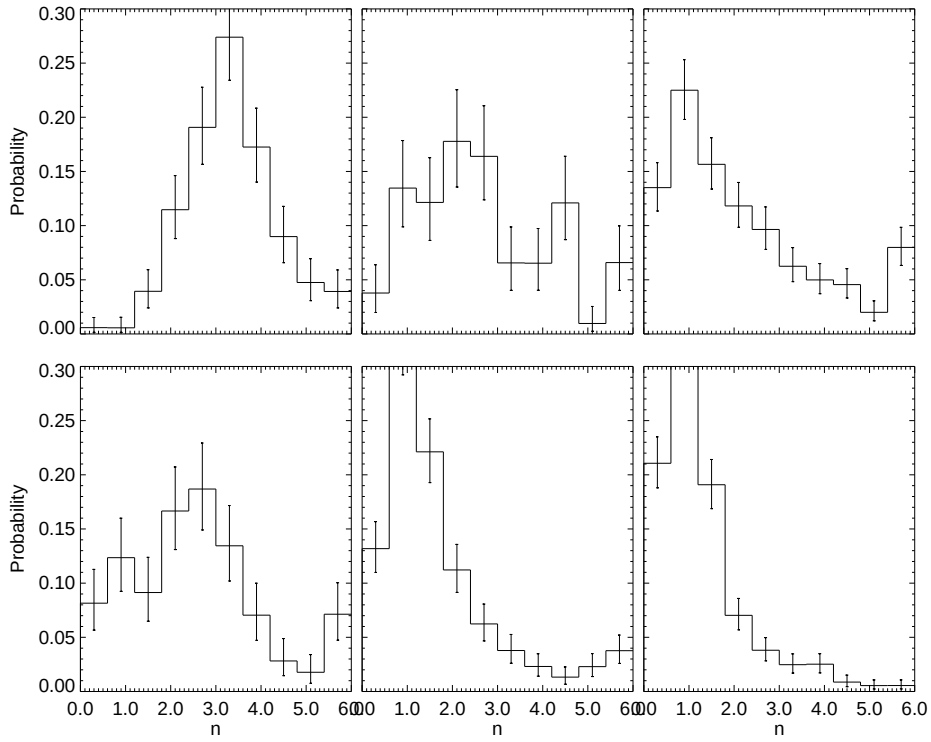


Figure 5.7: Similar to Fig.5.6 but showing the results from the population distribution for n from the 1DHBM for the most probable single component galaxies.

Table 5.2: Results for the one-component galaxies.

Low mass galaxies			
$0.02 < z < 0.06$			
Parameter	**Median**	**16%**	**84%**
R_e(/kpc)	3.01	2.35	4.17
n	2.36	1.37	3.14
$0.5 < z < 0.85$			
Parameter	**Median**	**16%**	**84%**
R_e(/kpc)	2.43	1.70	3.49
n	1.29	0.81	2.04
$0.85 < z < 1.25$			
Parameter	**Median**	**16%**	**84%**
R_e(/kpc)	2.28	1.69	3.12
n	1.01	0.66	1.59
High mass galaxies			
$0.02 < z < 0.06$			
Parameter	**Median**	**16%**	**84%**
R_e(/kpc)	4.52	3.68	5.79
n	3.26	2.62	3.88
$0.5 < z < 0.85$			
Parameter	**Median**	**16%**	**84%**
R_e(/kpc)	2.49	1.69	4.31
n	2.31	1.46	3.39
$0.85 < z < 1.25$			
Parameter	**Median**	**16%**	**84%**
R_e(/kpc)	2.27	1.56	3.50
n	1.56	0.85	2.71

In Figure 5.8, we summarise our results for the structural evolution of single component galaxies with the present-day mass of greater than (and equal to) or less than (and equal to) the mass of the Milky-way. To best compare with past literature we take the probability distributions from Figures 5.6 and 5.7 and randomly populate this parameter space weighting with the posterior probabilities. We have then take the median and 1-σ error regions from these distributions which best describes where a typical galaxy in some mass range will most likely lie at a specified redshift. We compare to several results from past literature. For our local Universe comparison group we have the median and 1-σ R_e for the elliptical galaxy sample from Gadotti (2009; grey open square) and local massive ($M_* > 10^{11} M_\odot$) galaxy sample from Szomoru et al. (2012; grey cross). We also compare to the Milky-way progenitor galaxies from van Dokkum et al. (2013; orange downward facing triangle) and $M_* = 10^{11.2} M_\odot$ progenitor galaxies taken from Patel et al. (2013; lime green right facing triangles).

The effective radii from both of mass samples share similarities at $z \sim 1$ with the most massive objects leading to slightly higher in upper percentile for the high mass sample. Since $z \sim 1$ the high mass sample have increased in size by a factor of ~ 1.8 whereas the low mass sample have only increased by a factor of ~ 1.2. The n values for both samples have greatly increased since $z \sim 1$, with the high mass sample having $n \sim 2.1$ at $z \sim 1$ and $n \sim 4.3$ at $z \sim 0$. The low mass sample have had a similar increase in n from $n \sim 1.5$ to 3.1. Thus both samples have had an increase in n by a factor of 2. van Dokkum et al. (2013) studied the evolving structures of Milky-way progenitors defined by following the same cumulative comoving number density over a redshift range of $0 < z < 3$ (see the dashed line in Fig.5.2 and 5.3. We see that our results are in agreement with theirs. However as they only fit each galaxy with a single Sérsic model, so two-components may be within their sample. This problem is the same when comparing to the high mass sample of Patel et al. (2013) but as they follow the progenitors of galaxies with masses $M_* = 10^{11.2} M_\odot$ (again see the dot-dot-dashed line in Fig. 5.2) we do not expect to have many two-component galaxies with such high masses in the local Universe. The results from Patel et al. (2013) show a larger growth in size than the median line for our high mass sample but it is also mentioned in Patel et al. (2013) that these high mass systems are very rare (for example, at a redshift range $1 < z < 1.5$ they only have 20 objects in their sample). The upper 1-σ distribution for our high mass sample extend more sharply, hinting that it is influenced by theses rare high mass systems following what was shown in Patel et al. (2013). As for the galaxies in the nearby Universe, are results match very well with the ellipticals from Gadotti (2009), which is not surprising since we are using the sample sample, even if we are using a different methodological approach.

We now explore at the joint relationship between the Sérsic parameters to investigate how the joint probabilities can help us explore structural evolution in a much more detailed way. Figure 5.9 shows the results from the 2DHBM for the Sérsic parameters of the single component galaxies. The contours encapsulate the probabilities calculated in the HBM process, where the red lines show the high mass sample in each redshift range and the black lines show the low mass objects. High mass objects are always brighter than their lower mass counter parts but only in the nearby Universe do they have a significant size difference. At $z \sim 0$ the two samples differ in location on the Kormandy relation but due to k-correction effects a direct comparison between redshift ranges would be incorrect. At $z \sim 1$ both mass groups have similar probability distributions with evidence that the high mass objects evolve more rapidly than the lower mass objects. We see positive correlations between the mass groups in the $R_e - n$ (Fig. 5.15) and $\mu_e - n$ relations of Figure 5.10 at $z \sim 0$. There is also evidence that low mass objects with more discy structures (i.e. $n < 2$) have a higher chance to lie

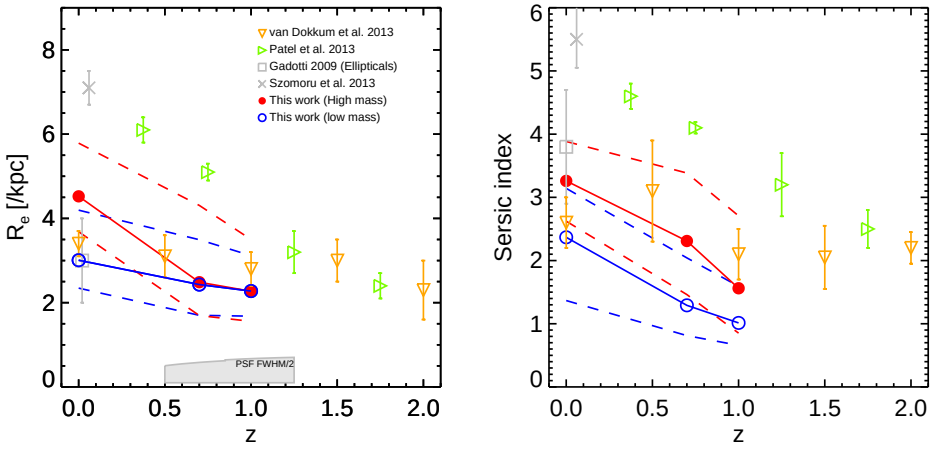


Figure 5.8: Effective radius and Sérsic index as a function of redshift of the one component galaxies, for our high mass sample (red filled circle) and low mass sample (blue circle) with the dashed lines showing the 1-σ error margins. We compare to the local elliptical sample from Gadotti (2009; grey open square), local massive ($M_* > 10^{11} M_\odot$) galaxy sample from Szomoru et al. (2012; grey cross), Milky-way progenitor galaxies from van Dokkum et al. (2013; orange downward facing triangle) and $M_* = 10^{11.2} M_\odot$ progenitor galaxies taken from Patel et al. (2013; line green right facing triangles). The grey box shows the PSF FWHM/2 limiting region over the high redshift range of our analysis.

off the positive relation seen for the more concentrated objects (i.e. $n > 2$).

Figure 5.11 compares directly how the two mass samples evolve over the redshift range by looking at the probability contours for each group separately and over plotting the previous redshift bin. In each panel, the red dashed contours compare how the population in the higher redshift bin is to the immediate lower redshift bin (black solid contours) for either the high mass (top row) or low mass (bottom row) samples. For both mass samples the probability distribution of n vs. R_e shows a negative correlation from $z \sim 1$ to $z \sim 0.7$ with lower n values having a higher chance that the effective radii will extend to larger values. At $z \sim 0.7$ the higher mass population appears to increase both in R_e and in n as we move from $z \sim 0.7$ to $z \sim 0$.

The lower mass populations follow a similar evolution to the high mass objects. There has been an insignificant change between $z \sim 1$ to $z \sim 0.7$ for the low mass sample, with the negative correlation in the $R_e - n$ plane still in place.The left panel shows a shift in the probability distributions between $z \sim 0.7$ to $z \sim 0$ we see that galaxies with lower n and higher R_e are still present in the local Universe, as well as $n \sim 2$ with smaller radii objects. We also see high R_e, high n population emerge between $z \sim 0.7$ and $z \sim 0$ resembling what was interpreted as a positive correlation in the $R_e - n$ plane.

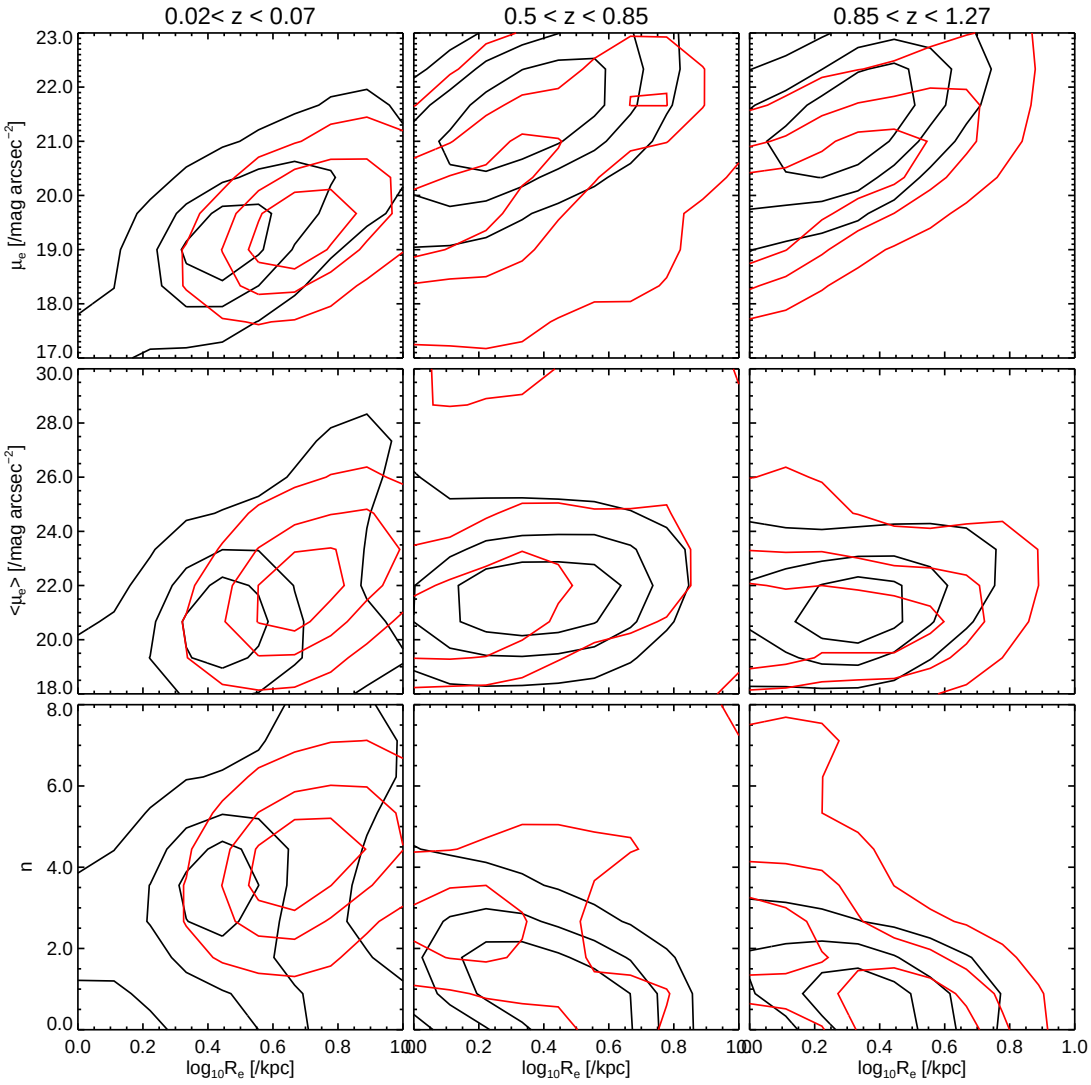


Figure 5.9: Results from the 2DHBM of the Sérsic parameters for the one component galaxies defined by the ΔBIC. From top to bottom: the effective surface brightness, the average effective surface brightness (without cosmological surface brightness dimming corrections) and n vs. R_e within the difference redshift ranges. The contour levels describe the most likely regions defined by the medians for each bin described by the 2DHBM. The red contours are for the high mass sample and the black contours are for the low mass sample. The contours show the 1, 2, and 3σ confidence regions.

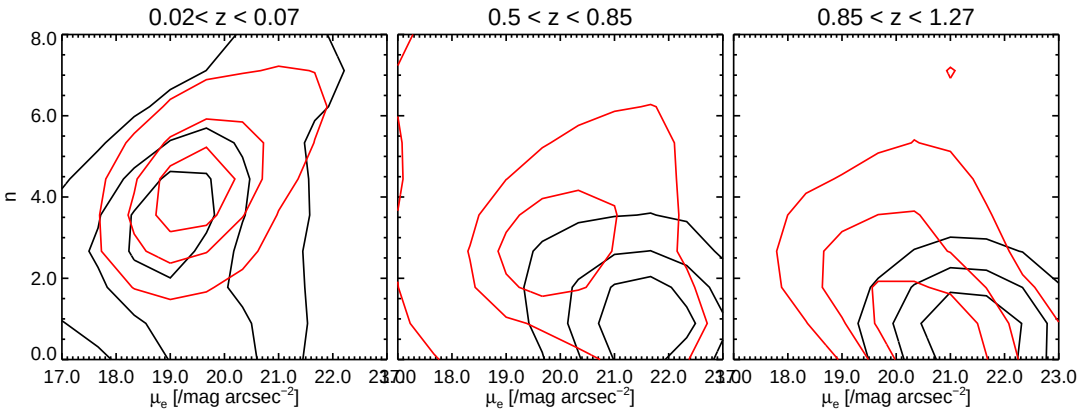


Figure 5.10: Similar to Fig.5.9 but showing the results from the 2DHBM $\mu_e - n$ plane for the single component systems over the redshift range $0 < z < 1.27$.

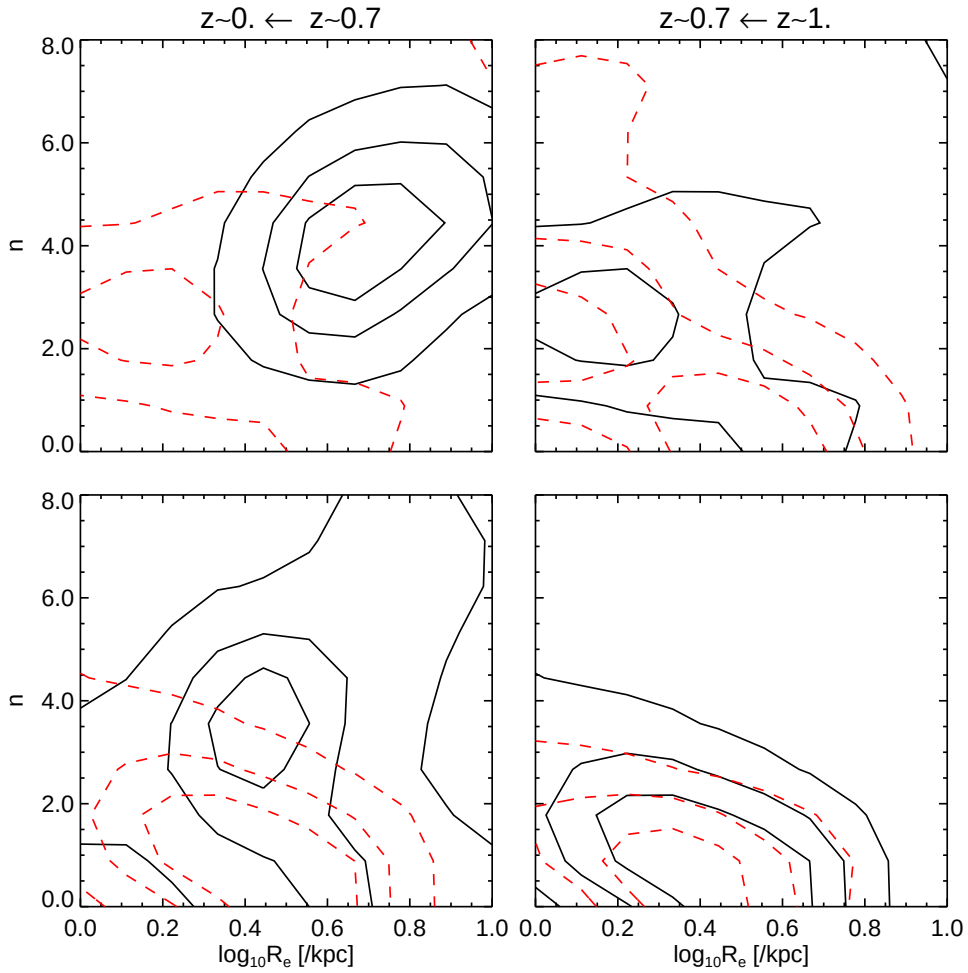


Figure 5.11: Probability contours for the $R_e - n$ plane from the 2DHBM for the single component galaxies. The top row shows the high mass sample and the lower row are the lower mass objects. We show how the populations evolve from $z \sim 1$ to $z \sim 0.7$ and $z \sim 0.7$ to $z \sim 0$. In each panel, the red dashed contours compares how the population at the higher redshift bin is to the immediate lower redshift bin (black solid contours).

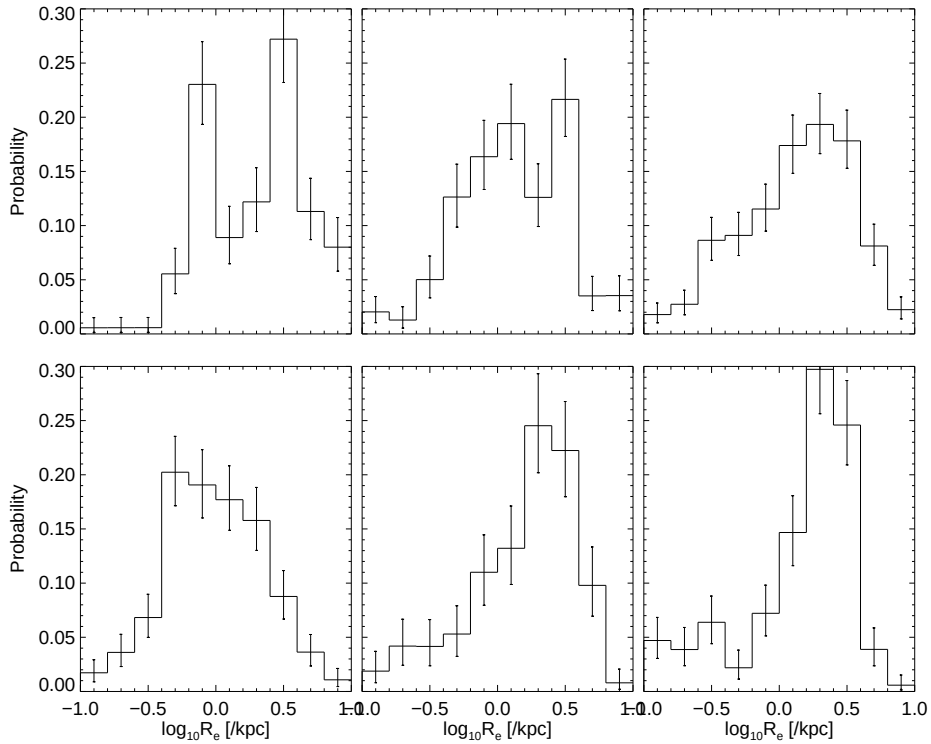


Figure 5.12: Similar to Fig.5.6 showing the results from the 1DHBM for the R_e of the most probable two-component galaxies defined using the ΔBIC. The top row shows results for the high mass sub-sample and the bottom row show the lower mass galaxy populations for all three redshift bins; $z \sim 0$, $z \sim 0.7$, and $z \sim 1$. The values of the histograms are the median values for the fractional counts for each bin estimated with the 1DHBM with the error bars showing the 16th and 84th percentiles.

5.6.2 two-component galaxies

In the previous sub-section we have shown that the single component galaxy population have evolved significantly since $z \sim 1$. Now we want to assess how the structure of the two-component galaxies has changed over this redshift range. We split this subsection to study the Sérsic (or bulge) and exponential (or disc) parameters separately and then look at how they can influence each other.

Sérsic parameter evolution

Fig. 5.12 shows the results from the 1DHBM for the population distribution of R_e for the two-component galaxies defined by the ΔBIC. At $z \sim 1$, the R_e distributions for the high mass (top row) and the low mass (bottom row) look very similar, with both probability distributions peaking at the higher radii end. We also find that the R_e distributions at $z \sim 0.7$ are again similar between the two mass populations. Only at $z \sim 0$ do we see a shift between the populations. The higher mass galaxies have split into two probability peaks at lower ($\sim$ 0.8kpc) and higher radii ($\sim$ 2.5kpc). The

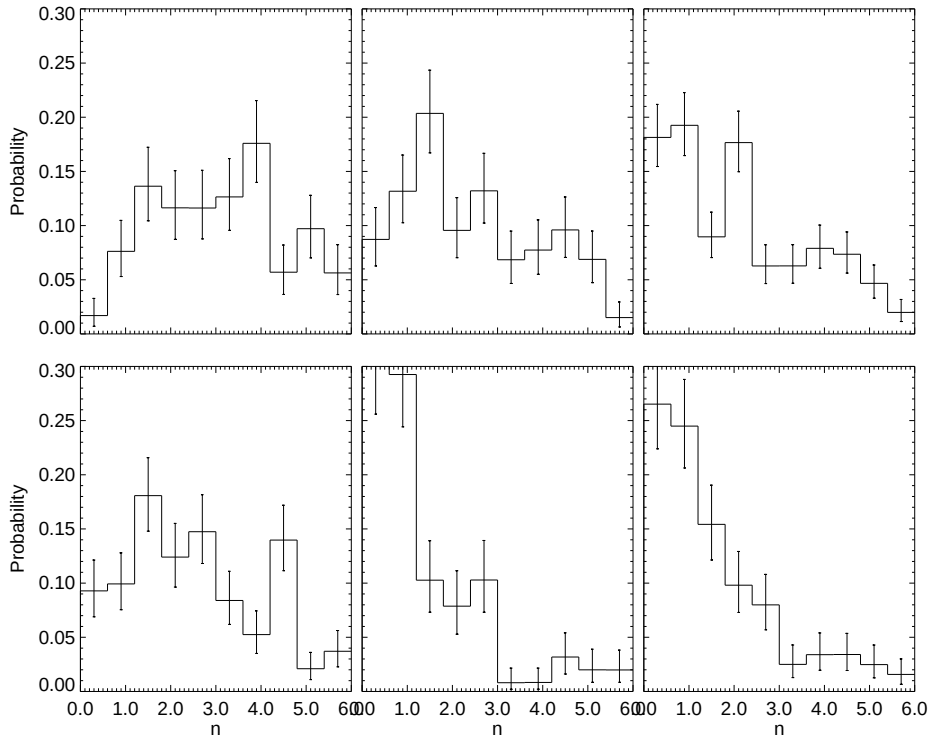


Figure 5.13: Similar to Fig.5.12 but showing the results from the population distribution for n from the 1DHBM for the most probable two-component galaxies.

lower mass populations show some decrease in size between $z \sim 0.7$ and $z \sim 0$.

Figure 5.13 presents the results for the 1DHBM of the n values for the population of two-component galaxies. At $z \sim 0$ we see clear differences between the two mass groups, where the high mass sample are more likely to have higher n values (peaking around $n \sim 3.5$) than the lower mass population where the probability distribution peaks around $n \sim 1.5$. At $z \sim 1$ both mass groups have similar probability distributions for n with the higher mass population extending to slightly higher values. Since $z \sim 1$ both populations show signs of an evolving shape profile which can be interpreted as bulges becoming more compact over time, but this evolution is again stronger for the high-mass bulges.

In Figure 5.14, we summarise our results for the structural evolution of the two compoent galaxies. We perform a similar analysis as before with the single component galaxies, where we use the 1DHBM results for R_e and n and re-populate the parameter space weighted with the predicted probability distributions. We then take median and 16th and 18th percentile values from these predictions. We then compare to several results from the literature. We show the results from Gadotti (2009) for his classical (closed grey square) and pseudo-bulges (open grey square). We also show the results

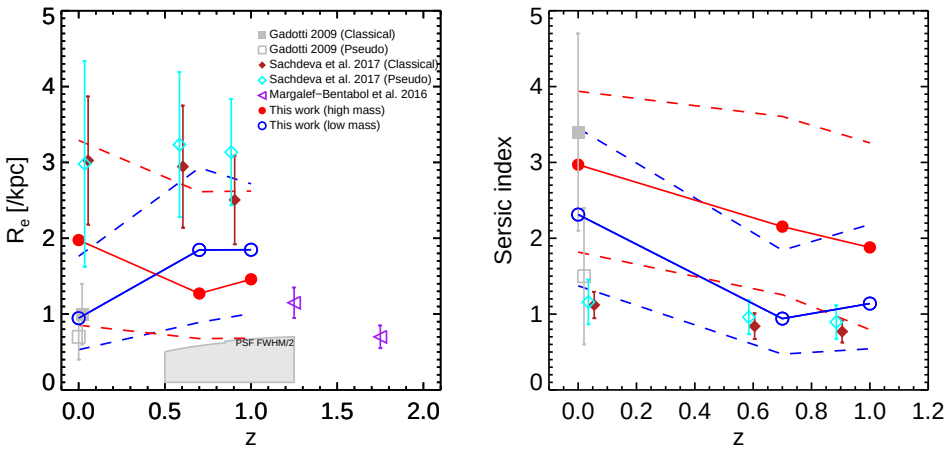


Figure 5.14: Effective radius and Sérsic index as a function of redshift of the two-component galaxies, for our high mass sample (red filled circle) and low mass sample (blue circle) with the dashed lines showing the 1-σ error margins. We compare to the local sample from Gadotti (2009; pseudo-bulges is the grey open square and classical bulges is the closed grey square). The cyan open diamonds and dark red closed diamonds show the pseudo and classical bulges from Sachdeva, Saha and Sihgh (2017) respectively. The purple left-facing triangles show the results from Margalef-Bentabol et al. (2016) for the bulges of there two-component galaxies. The grey box shows the PSF FWHM/2 limiting region over the high redshift range of our analysis.

from Sachdeva et al. (2017) where a magnitude limited sample of galaxies from GOODS-South ($0.4 < z < 1.0$) and SDSS ($0.02 < z < 0.05$) were separated into classical and pseudo-bulges. We also include the bestfit R_e values for two-component galaxies from Margalef-Bentabol et al. (2016) of their R_e best fit values for two-component galaxies (left-facing purple triangles). Both Gadotti (2009) and Margalef-Bentabol et al. (2016) both did 2D photometric fits with the codes BUDDA (de Souza et al., 2004) and GALFIT (Peng et al., 2002) (GALAPAGOS) respectively, whereas Sachdeva et al. (2017) performed 1D photometric fits which could lead to overall biases in their results (as discussed in Section 2.1).

Focusing first on the high mass population, the growth in R_e is very subtle (a factor of ~ 1.3) from $z \sim 1$ to $z \sim 0$. While the low mass sample appears to have an overall decrease in size with bulges being roughly a factor ~ 0.5 smaller than they were at $z \sim 1$. The low mass sample also appear to have slightly higher radii at $z \sim 1$. Our low mass sample agree with both the classical- and pseudo-bulge R_e estimates from Gadotti (2009). The results from Sachdeva, Saha and Sihgh (2017) are systemically larger than our estimates but this could be due to the biases when fitting in only one-dimension. For both mass populations the median n values increase in size with factors ~ 2 and ~ 1.6 for the low mass and the high mass samples respectively. The high mass populations appear to always have higher n values that grow at an almost constant rate, whereas the low mass galaxies have a somewhat constant n median value between $z \sim 1$ and $z \sim 0.7$ with the change

Table 5.3: Results for the two-component galaxies.

Low mass galaxies			
$0.02 < z < 0.06$			
Parameter	**Median**	**16%**	**84%**
R_e(/kpc)	0.95	0.53	1.78
n	2.31	1.37	3.45
h(/kpc)	5.41	3.16	10.74
B/T	0.188	0.085	0.51
$0.5 < z < 0.85$			
Parameter	**Median**	**16%**	**84%**
R_e(/kpc)	1.85	0.314	2.92
n	0.94	0.47	1.84
h(/kpc)	7.67	4.19	13.46
B/T	0.30	0.11	0.42
$0.85 < z < 1.25$			
Parameter	**Median**	**16%**	**84%**
R_e(/kpc)	1.85	1.01	2.72
n	1.14	0.54	2.18
h(/kpc)	4.76	2.89	8.75
B/T	0.33	0.10	0.53
High mass galaxies			
$0.02 < z < 0.06$			
Parameter	**Median**	**16%**	**84%**
R_e(/kpc)	1.98	0.85	3.29
n	2.97	1.82	3.94
h(/kpc)	13.61	8.17	22.18
B/T	0.26	0.11	0.66
$0.5 < z < 0.85$			
Parameter	**Median**	**16%**	**84%**
R_e(/kpc)	1.27	0.68	2.61
n	2.15	1.26	3.61
h(/kpc)	8.97	4.58	18.58
B/T	0.25	0.12	0.47
$0.85 < z < 1.25$			
Parameter	**Median**	**16%**	**84%**
R_e(/kpc)	1.46	0.68	2.62
n	1.88	0.79	3.26
h(/kpc)	10.12	5.14	19.01
B/T	0.25	0.10	0.47

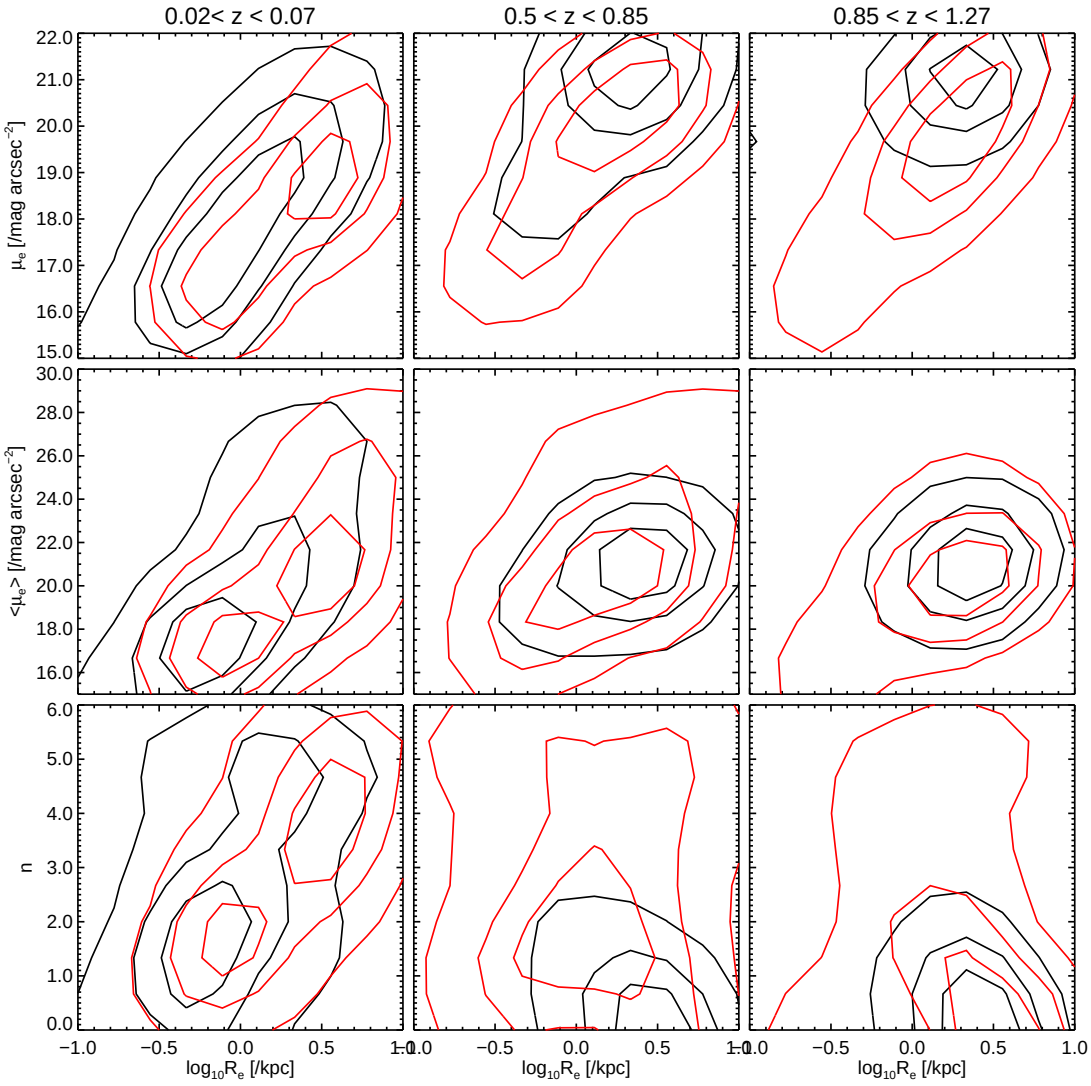


Figure 5.15: The 2DHBM of the Sérsic parameters for the two-component galaxies defined by the ΔBIC. From top to bottom: the effective surface brightness, the average surface brightness and n verses R_e within the different redshift ranges. The contour levels describe the most likely regions defined by the medians for each bin described by the 2DHBM. The red contours are for the high mass sample and the black contours are for the low mass sample.

happening between $z \sim 0.7$ and $z \sim 0$. Our results are consistent with those from Gadotti (2009) but due to the over estimates in R_e we see underestimates in n from Sachdeva, Saha and Sihgh (2017).

Figure 5.15 shows the joint probability distributions for the Sérsic parameters of the two-component galaxies from the 2DHBM. There are again differences between the high mass and low mass populations at $z \sim 0$. Both populations share a similar trend in $\mu_e - R_e$ with the higher mass systems having a higher probability to be more extended and dimmer. These relations are also present in

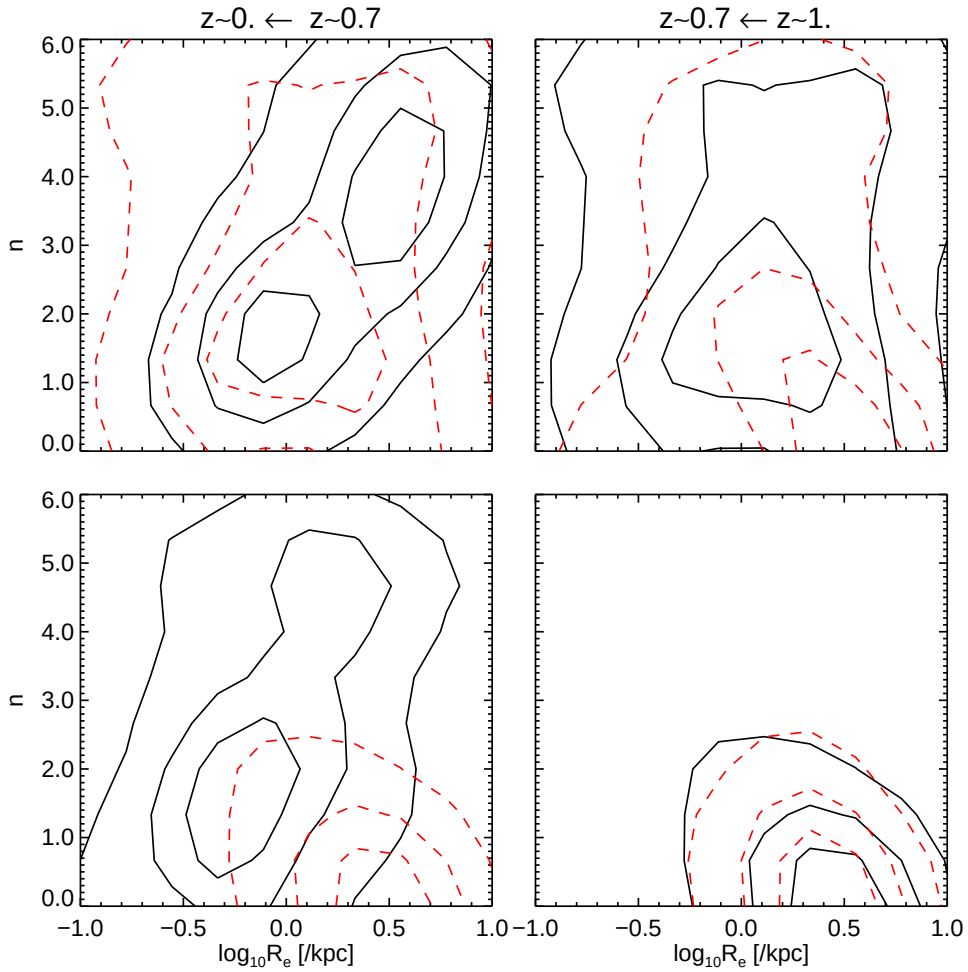


Figure 5.16: Similar to Fig.5.11 showing the probability contours for the $R_e - n$ plane from the 2DHBM for the two-component galaxies. The top row shows the high mass sample and the lower row are the lower mass objects. We show how the populations evolve from $z \sim 1$ to $z \sim 0.7$ and $z \sim 0.7$ to $z \sim 0$. In each panel, the red dashed contours compares how the population at the higher redshift bin is to the immediate lower redshift bin (black solid contours).

the higher redshift bins. In $< \mu_e >$ vs R_e at $z \sim 0$, we observe both mass samples to have the same distribution of μ_e at low R_e, but a different slope in the relation as we increase in radius. Galaxies with smaller radii are more likely to have more discy structures (for both mass groups) as seen in the $R_e - n$ plane. The higher mass population are more likely to have larger radii than the lower mass population but the overall trend between R_e and n is positive in both mass samples. At $z \sim 1$, we observe a similar swap in the correlation between R_e and n as we did for the single component galaxies. The two mass populations are very similar,but only the high mass samples extend into the high n region. At $z \sim 0.7$ the two populations differ much more as we begin to see some evolutionary differences between the populations.

Figure 5.16 compares directly how the two mass samples evolve since $z \sim 1$ by looking at the

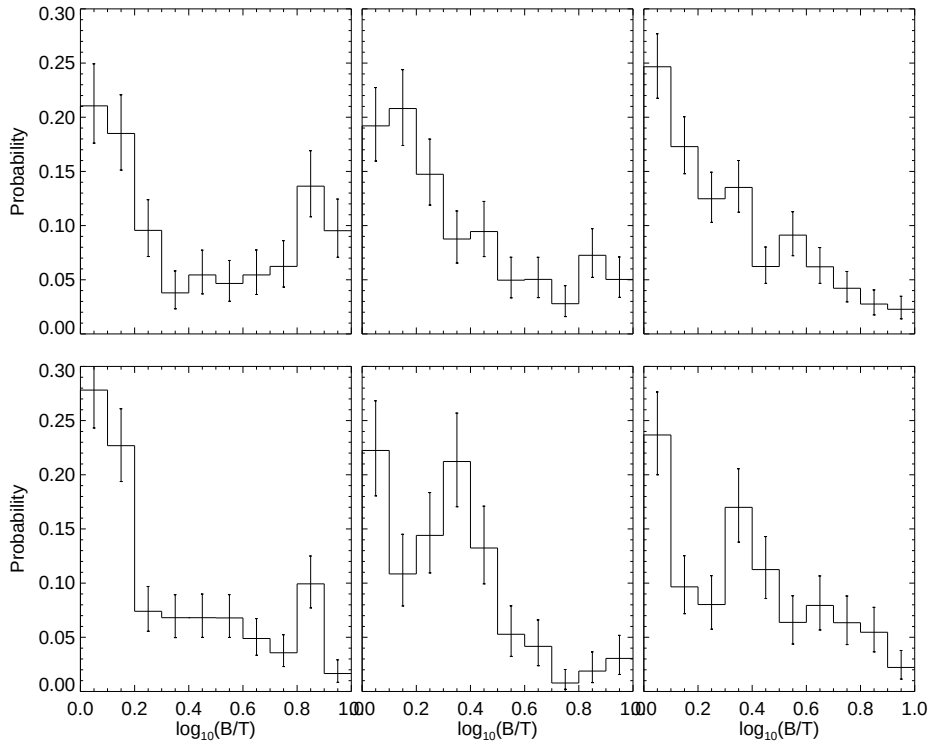


Figure 5.17: Similar to Fig.5.12 showing the results from the 1DHBM for the B/T probability distribution of the two-component galaxies.

probability contours for each group separately and over plotting the previous redshift bin to look for similarities and differences. At $z \sim 1$ both mass populations have a higher probability to be large and a more discy in structure, with the smaller radii objects most likely being more concentrated. For the high mass sample, at $z \sim 1$ there are already galaxies with large n and R_e and as we move to $z \sim 0.7$ the entire population shifts onto a positive relation in the $R_e - n$ plane just as at $z \sim 0$. From $z \sim 1$ to $z \sim 0.7$, the low mass sample show little sign of change, keeping the negative correlation in the $R_e - n$ plane. However, the contours at $z \sim 0.7$ do look to extend into the most likely region at $z \sim 0$. For the lower mass population the most significant changes happen between $z \sim 0.7$ and $z \sim 0$, as the correction shifts from positive to negative with a push into the $n > 3$ region.

The bulge-to-total ratio, B/T, is tightly related to the luminosity or mass ratio of the components of a galaxy. It is thought that most properties of galaxies are related to this ratio. Thus it is key to understand how the B/T ratio is established across time. In Figure 5.17 we plot the 1DHBM for the B/T ratio for the two-component galaxies with the top row for the high mass sample and the bottom row for the low mass sample. Since $z \sim 1$ galaxies with $B/T < 0.3$ have been dominant. The contribution from intermediate $0.3 < B/T < 0.7$ has decreased since $z \sim 1$. In the same time frame, there appears to be a growing probability for the most bulge dominated galaxies ($B/T > 0.7$).

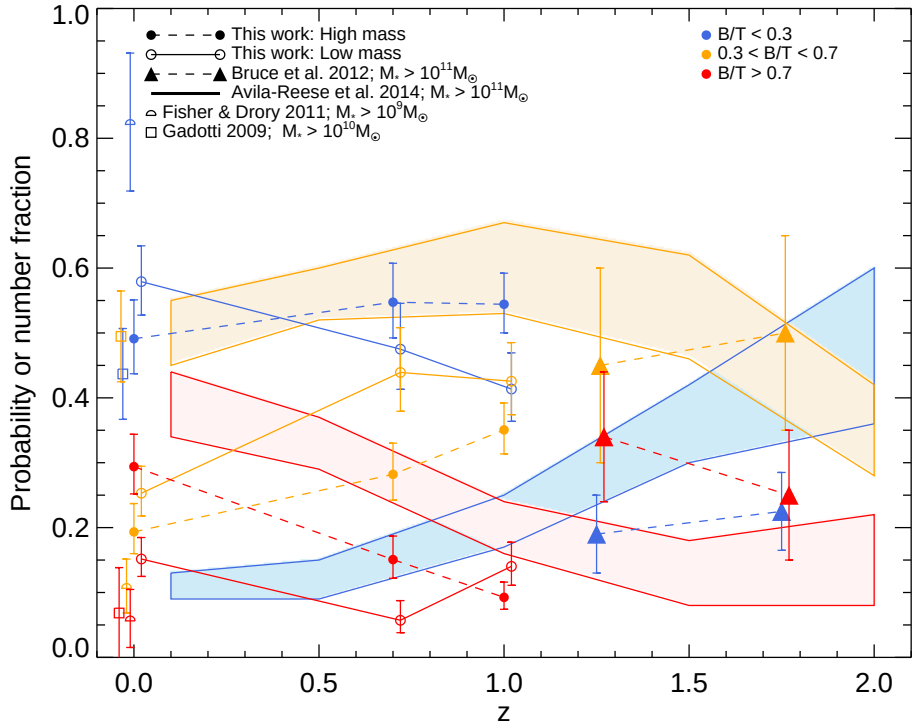


Figure 5.18: Probability of the high mass (filled circle-dashed line) and low mass (circle-solid line) populations according to their B/T values (indicated by the colour code in the upper right legend) as a function of z. Also shown is the number fractions from the study by Bruce et al. (2012) for galaxies with $M_* > 10^{11} M_\odot$ (triangle-dashed line). The open squares indicate the full sample from Gadotti (2009) and the half-circles show the results from Fisher & Drory 2011 ($M_* > 10^9 M_\odot$). The filled regions indicated the number fraction of the semi-empirical galaxies more massive than $10^{11} M_\odot$ from Avila-Reese et al. 2014. The contrasting mass rages make direct comparisons difficult, although the trends do indicate the importance of mass in galaxy evolutionary scenarios.

In figure 5.18 we plot the evolution of the integrated probability distributions for three B/T bins. The high mass sample is shown by the filled circles and lower mass sample is shown by the open circles, respectively. We compare to two $z \sim 0$ works, the first is that from Gadotti (2009; open squares) of whom did bulge-disc decompositions on ~ 1000 SDSS galaxies (see previous chapter for details) and the second is the works from Fisher & Drory (2011) (open half circles) on 99 galaxies with $M_* > 10^9 M_\odot$ in the local 11 Mpc volume. For the local Universe the most massive galaxies have a higher probability to be bulge dominated and with the lower mass galaxies showing that lower B/T values dominate. We compare our results to the high redshift ($1 < z < 2$) number fractions reported for high mass galaxies ($M_* > 10^{11} M_\odot$) in Bruce et al. (2012; solid triangles linked by the dashed lines): $B/T \leq 0.3$ (blue), $0.3 < B/T \leq 0.7$ (orange), and $B/T > 0.7$ (red). To aid our understanding of the evolutionary paths of two-component galaxies add the blocked regions indicating the number fractions (Poissonian errors on the number counts) predicted for the semi-empirical $M_* > 10^{11} M_\odot$ galaxies selected from the cosmological simulations presented in Avila-Reese et al. (2014).

The morphological mix of the observed massive galaxies in our analysis appears to disagree with Bruce et al. (2012). Who find both low and high B/T to be less common than intermediate values at $z > 1$. We find that this population has dominated since $z \sim 1$ with a decrease in the population from $z \sim 0.7$. For systems with $B/T > 0.7$, Bruce et al. (2012) reports the start of a growth period for these systems at $z \sim 1$, we also show that from $z \sim 1$ these systems increase in probability but to a lesser extent. As for the intermediate B/T systems ($0.3 < B/T \leq 0.7$), we do see a flow from the works presented in Bruce et al. (2012) as these systems decrease in probability (or number) over time. Comparing to the semi-empirical massive galaxies we again find similarities and differences. We disagree with the evolutionary trends of the $B/T \leq 0.3$ systems predicted by Avila-Reese et al. (2014) where by $z \sim 0$ these disc dominated systems are rarer than the bulge dominated galaxies. Since $z \sim 1$, fraction of semi-empirical galaxies with $B/T > 0.7$ increases, which is mirrored by our massive galaxy sample with $B/T > 0.7$ probabilities. The decrease in $0.3 < B/T \leq 0.7$ semi-empirical galaxies since $z \sim 1$ again matches how our intermediate B/T systems evolve. Our observations and the simulations follow similar trends but are offset from the semi-empirical galaxies by factors of ~ 2 and ~ 1.3 for the $B/T > 0.7$ and $0.3 < B/T \leq 0.7$ systems respectively.

Figure 5.19 shows the probability contours from the 2DHBM for the $B/T - R_e$ plane as we move from $z \sim 1$ to $z \sim 0.7$ and $z \sim 0.7$ to $z \sim 0$ (top row is for the high mass and bottom row for the lower masses). For both mass sub-samples there appears to be insignificant change between $z \sim 1$

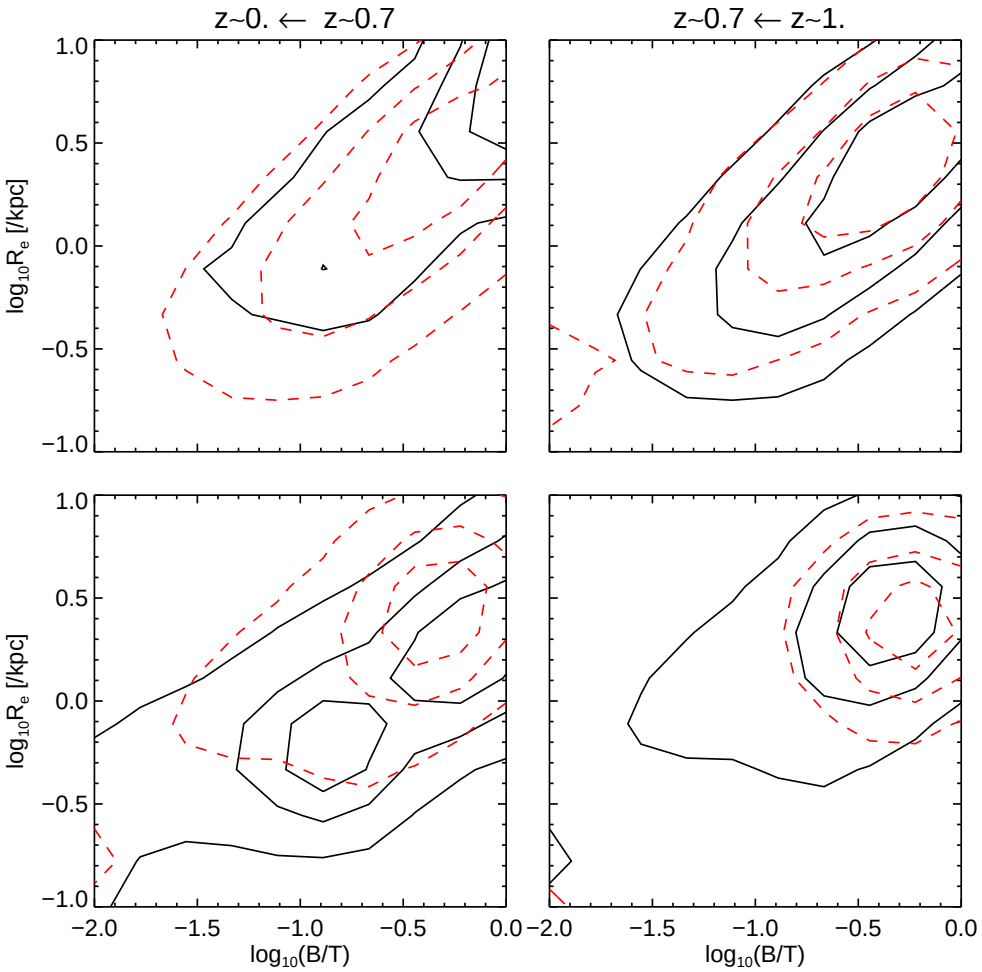


Figure 5.19: Similar to Fig.5.16 but showing the probability contours for the $B/T - R_e$ plane from the 2DHBM for the two-component galaxies.

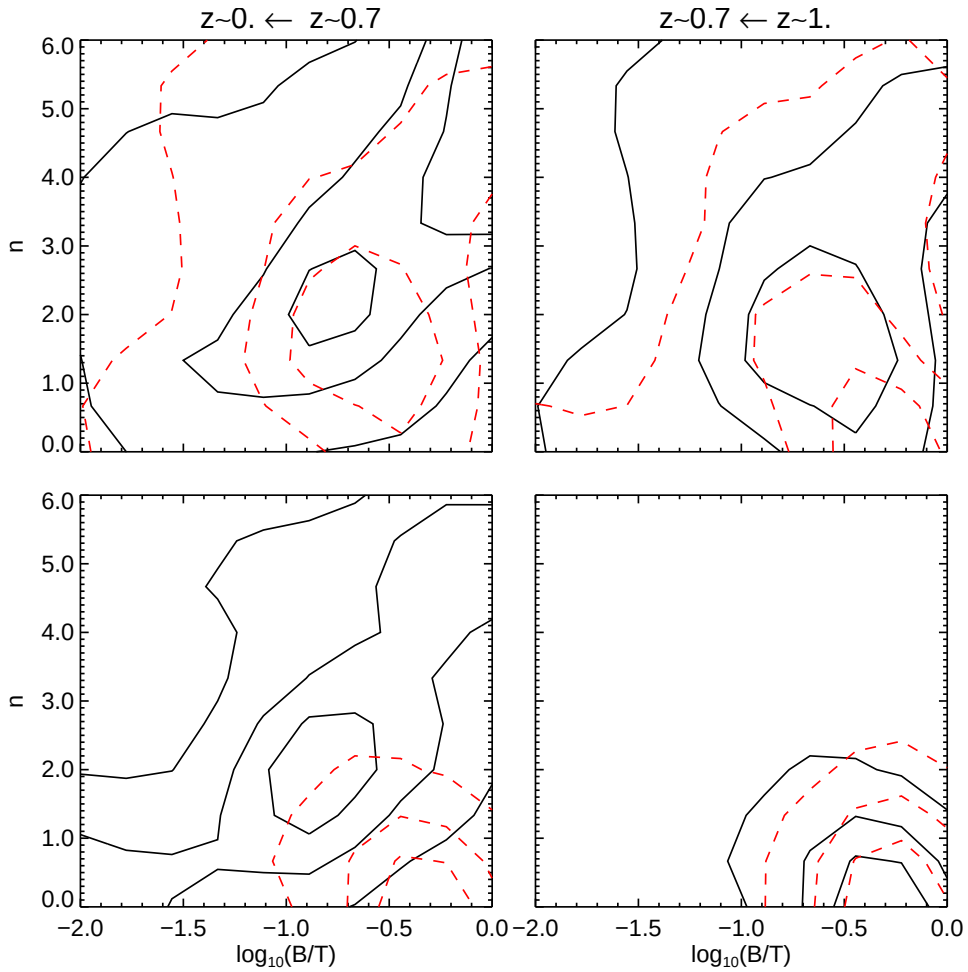


Figure 5.20: Similar to Fig.5.16 but showing the probability contours for the $B/T - n$ plane from the 2DHBM for the two-component galaxies.

to $z \sim 0.7$ with the more bulge dominated galaxies having larger radii. From $z \sim 0.7$ to $z \sim 0$ the high mass sample becomes increasingly more likely to be bulge dominated. We see here again the shrinking in R_e for the lower mass sample but the overall correlation in $B/T - R_e$ remains. For lower masses there appears to be two emergent probable locations for galaxies to lie in either: larger B/T and R_e or small B/T and R_e at $z \sim 0$.

Figure 5.20 shows the $B/T - n$ plane from the 2DHBM analysis of the two-component galaxies. The probability contours show more of a diversity than previously seen. The low mass and high mass populations have some similarities in that there is evidence that at $z \sim 1$ discy bulges are dominant. The changes between $z \sim 1$ to $z \sim 0.7$ for the lower mass population with discy bulge-dominate galaxies persist. We see that the lower mass population changes most significantly between $z \sim 0.7$ to $z \sim 0$. The discy bulges move up in n and down in B/T similar to the high mass sample with the emergent probability of higher n and B/T values. For the high mass sample there already is a population of bulges with $n > 2$ that are dominant and as we move to $z \sim 0.7$ the probability of this population increases. We see by $z \sim 0$, galaxies within the higher mass sample have fallen onto a positive relation in the $B/T - n$ plane with bulges with $n > 3$ having higher B/T values. There is however evidence that the once dominant discy bulge population is still present in the local universe but has since moved up in n and down in B/T.

Exponential parameter evolution

We have so far shown how the bulges of our two-component galaxies have grown since $z \sim 1$. We now look at the structures of their disc counter-parts to assess any evolutionary trends. Figure 5.21 shows the 1DHBM for the disc scale length for the two-component galaxies (again the top row is for the higher mass sample and the bottom row is for the low mass sample). Structural changes for the size of the disc are much more subtle than those of their bulges. The higher mass sample have had almost a constant disc size since redshift $z \sim 1$ (median $h \sim 10kpc$) with a slight increase in size from $z \sim 0.7$ to $z \sim 0$. The lower mass population show a decrease in size from $z \sim 0.7$ to $z \sim 0$ (with the median decreasing by a factor of ~ 1.3). Between $z \sim 1$ to 0.7 the probability distributions are similar, with slightly smaller disc sizes compared to the higher mass populations.

Figure 5.22 shows that the strong correlation between the central surface brightness and the disc scale length persists back to $z \sim 1$. In each redhsift bin the higher mass population (red contours) show a shifted $\mu_0 - h$ relation towards the larger brighter end. Again it is observed that the discs of both mass populations show little change between $z \sim 1$ to 0.7 with discs evolving more significantly

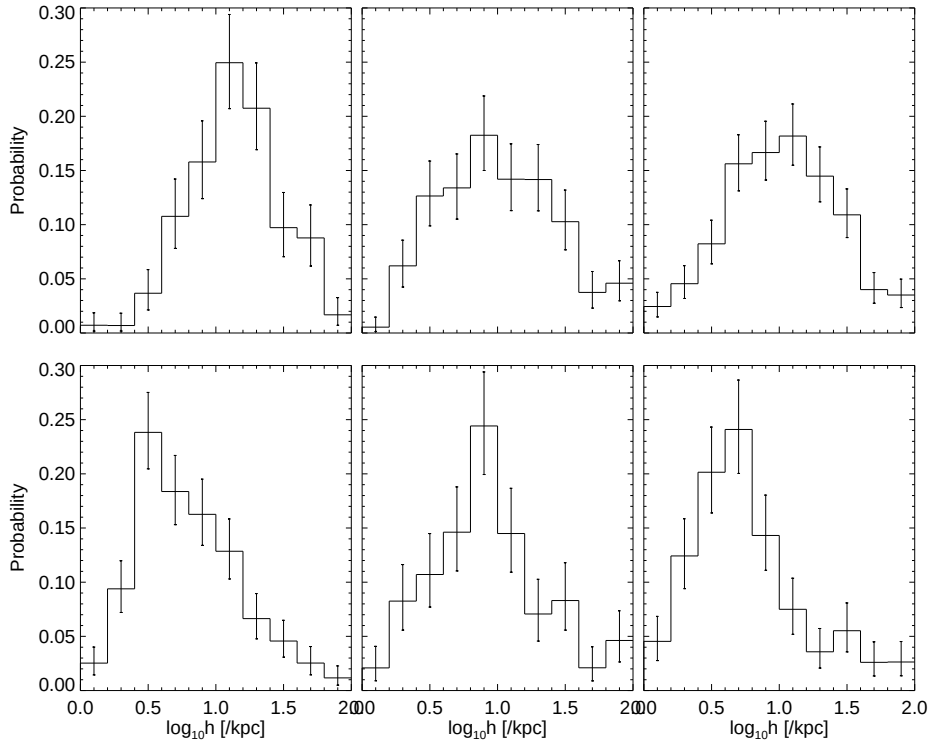


Figure 5.21: Similar to Fig.5.12 showing the results from the 1DHBM for the h probability distribution of the two-component galaxies.

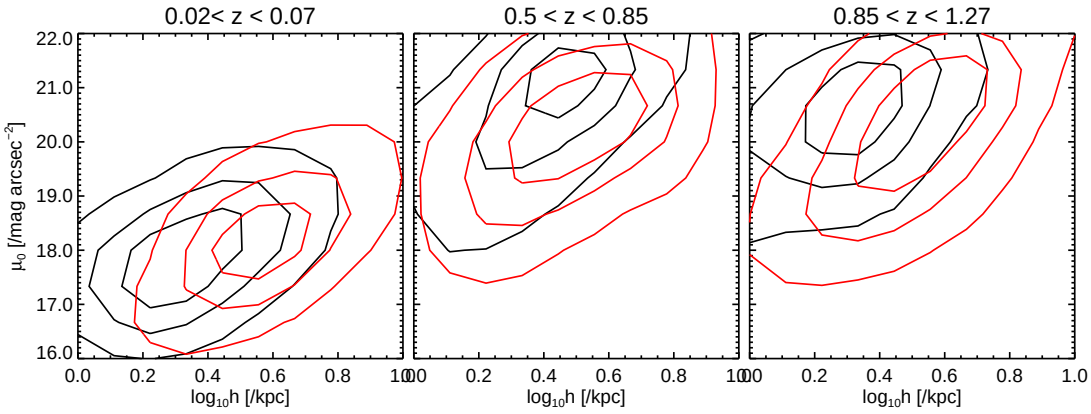


Figure 5.22: Similar to Fig.5.15 but showing the probability contours for the $\mu_0 - h$ plane from the 2DHBM for the two-component galaxies. The red contours are show the high mass sample and the black contours are for the low mass sample.

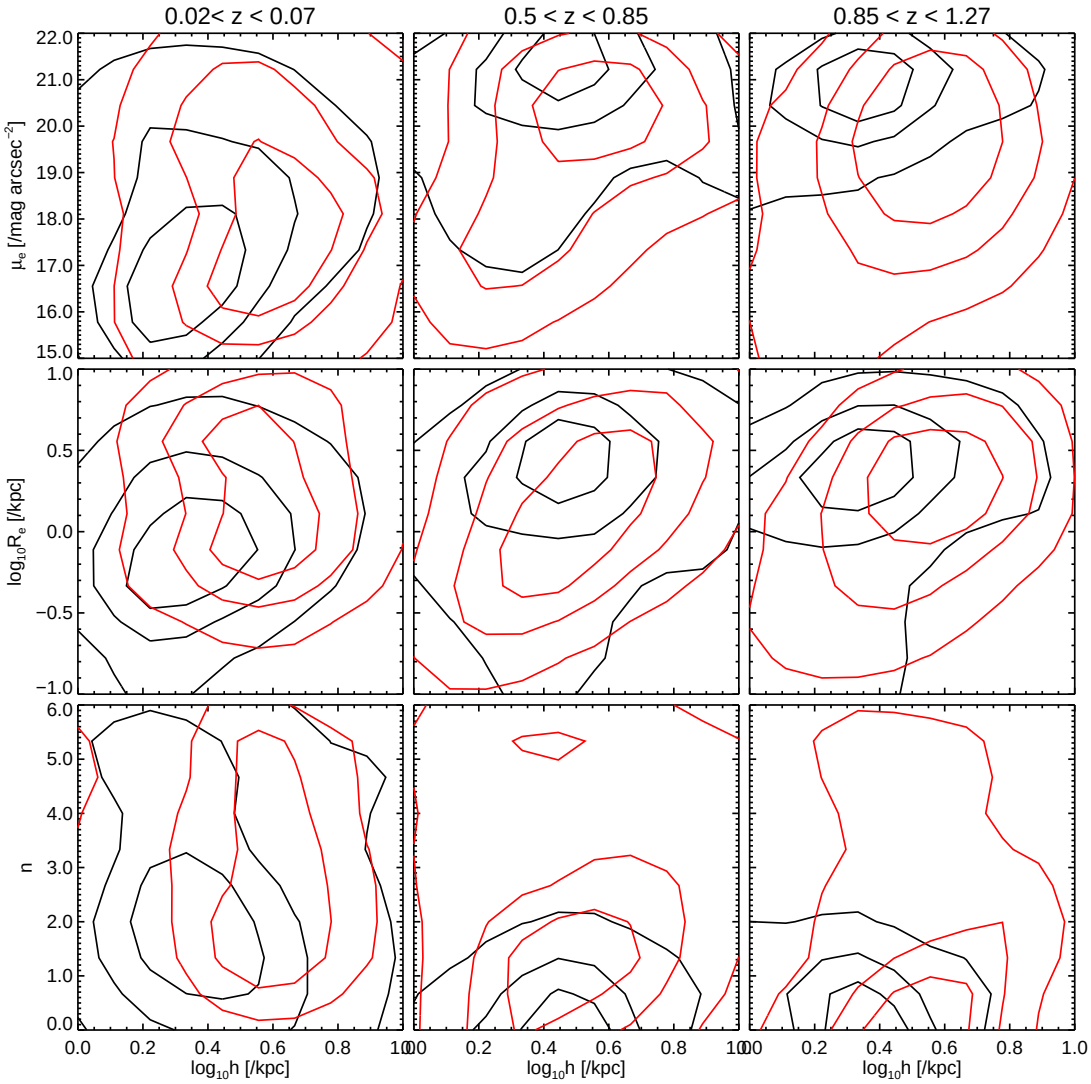


Figure 5.23: Similar to Fig.5.16 but showing the probability contours of h verses the Sérsic parameters from the 2DHBM for the two-component galaxies.

from $z \sim 0.7$ to $z \sim 0$.

Sérsic-Exponential parameter evolution

We have observed that the components that make up the two-component galaxies have followed distinct evolutionary paths of their own but how have they interacted with each other over the same time frame?

In Figure 5.23 we show the joint probability distributions of the disc scale lengths vs. the Sérsic parameters for the bulge components from the 2DHBM. Correlations between h and the Sérsic parameters are generally weak for both mass populations at all redshifts. Larger discs in the high mass

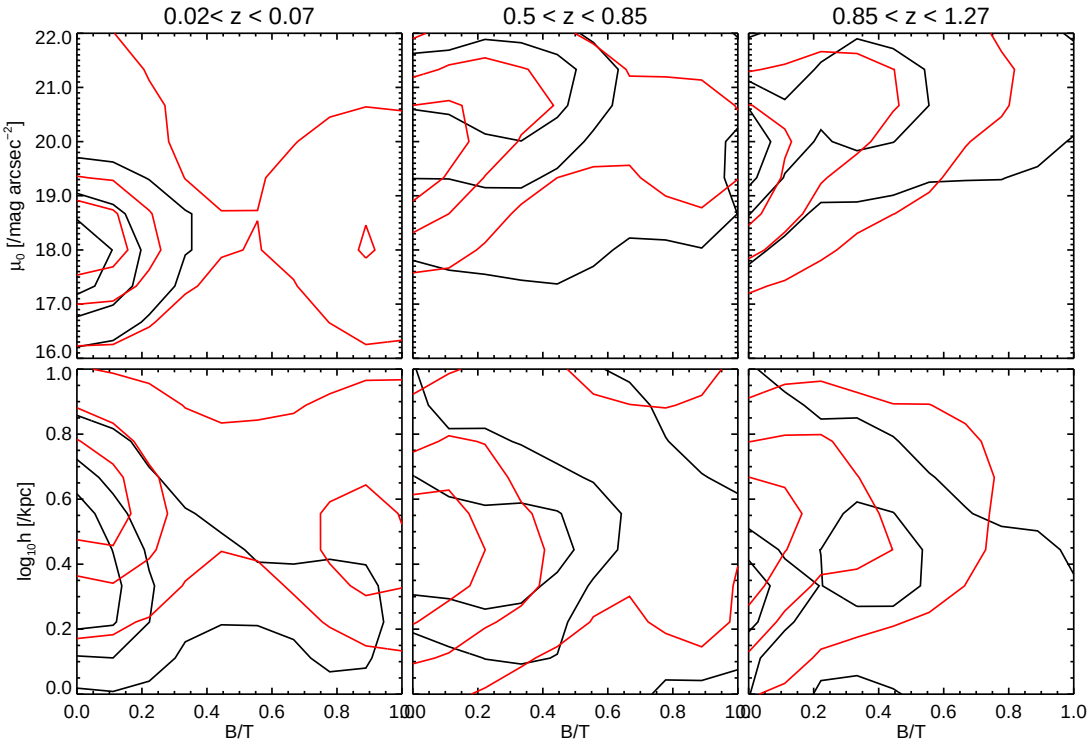


Figure 5.24: Similar to Fig.5.16 but showing the probability contours of B/T verses the exponential parameters from the 2DHBM for the two-component galaxies. The red and black contours are for the high mass and low mass populations respectively.

sample do tend to have bigger, dimmer bulges at $z \sim 0$ although for the same disc size the bulges have a probability distribution that span from discy to concentrated n values. The opposite can be said for the low mass sample where discs are more likely to be smaller with smaller brighter and more discy bulges. At $z \sim 0.7$ the two mass samples in the $R_e - h$ plane share the same probability space with both distributions following a positive relation. Within the same redshift bin, the $n - h$ plane for both populations falls down to lower n values with the lower and higher mass sample moving down and up in h, respectively. In the highest redshift bin, we find insignificant change in the relationships between h and the Sérsic parameters.

We can also observe how the dominance of the bulge component effects the disc parameters as a function of B/T. Figure 5.24 presents the probability distributions for the $B/T - \mu_0$ and $B/T - h$ planes for the two-component galaxies. There appears to be no correlation between B/T and the exponential parameters. We show that at $z \sim 1$ galaxies in both samples have a similar parameter probability distributions, and between $z \sim 1$ and 0.7, high B/T galaxies begin to emerge. Between $z \sim 0.7$ and $z \sim 0$ a bimodality in the $B/T - h$ plane for the high mass sample is observed with the lower mass population reverting back to a more disc dominated population.

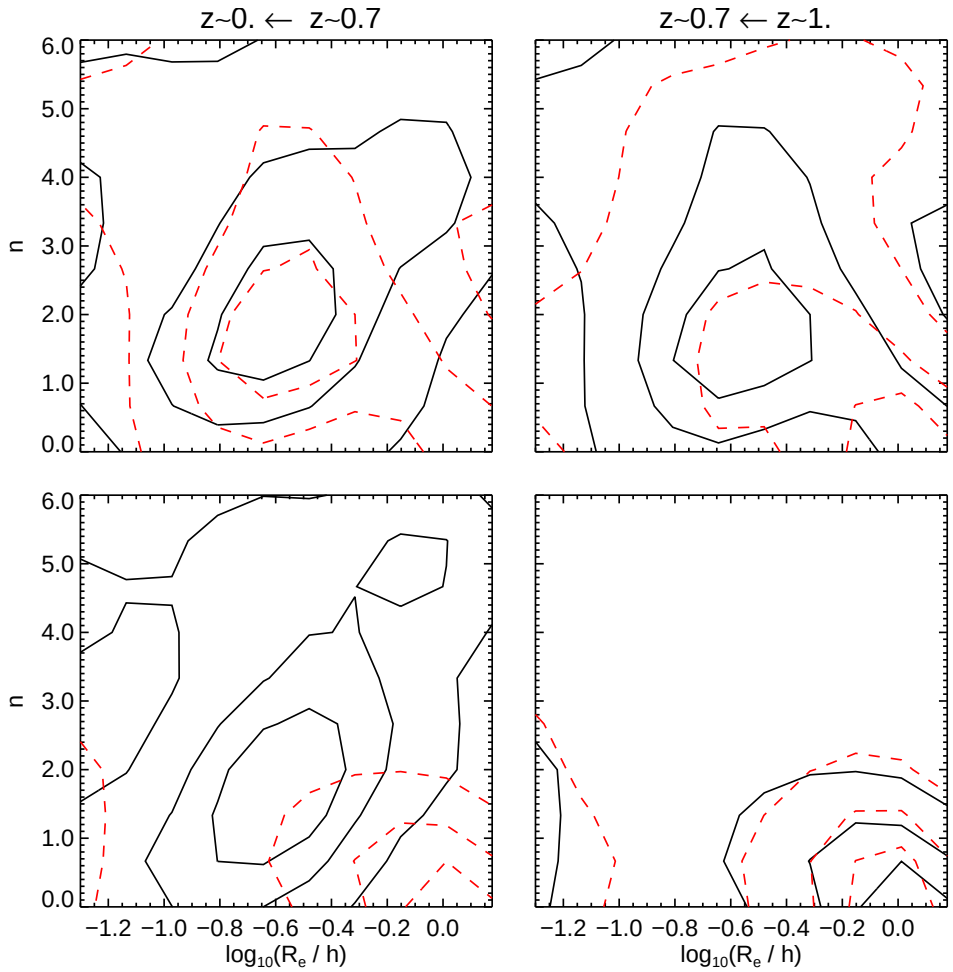


Figure 5.25: Similar to Fig.5.16 but showing the probability contours of the ratio between R_e and h verses n from the 2DHBM for the two-component galaxies.

In Figure 5.25 we look at the relation between the ratio of R_e and h to n and how it has evolved since $z \sim 1$. In Figure 5.20 we showed that B/T is correlated with n at $z \sim 0$, so we expect to find a similar correlation in the $R_e/h - n$ plane. The $R_e/h - n$ describes how the size parameters of both components and n are directly related. Massive galaxies in the local Universe can be seen to have similar probability distributions to their progenitors at $z \sim 0.7$ where a positive correlation is seen. For the same mass group, there is some evidence that galaxies with high n and a larger R_e/h ratio are also present at $z \sim 1$. Although there is a higher probability that these galaxies will have a larger R_e/h with $n < 2$. In the local Universe, the lower mass galaxies follow a similar positive relation to those with higher masses. However, the higher mass sample appear to evolve more rapidly. At $z \sim 0.7$ and $z \sim 1$ lower mass two-component galaxies are more likely to have $n < 2$ and a larger R_e/h ratio.

5.7 Discussion and conclusions

In this chapter we studied the structural evolution of 1368 COSMOS/3D-HST galaxies in the redshift range $0.5 \leq z \leq 1.27$, and compared them to a local sample of 454 SDSS galaxies. We have applied a HBM including the posterior probabilities from 2D photometric bulge-disc decompositions to achieve a full probabilistic representation of galaxy structures. The non-parametric nature of our HBM means we can re-populate the parameter space based on the population probability function. We have then used this to compare directly with previous observational and theoretical studies. In the light of these new results, we now discuss the importance of our findings in relation to theoretical evolutionary scenarios. We first summarise the method and list our major finding, then we discuss the implications:

I. We used stellar population models to obtain an estimate of the progenitors of galaxies with masses greater than that of the Milky-Way ($M_* > 10^{10.66} M_\odot$) in the local Universe and the progenitors of $10^{10} < M_* \leq 10^{10.66} M_\odot$ galaxies at $z \sim 0$. To properly compare our local sample ($0.04 < z < 0.06$) in the SDSS i-band to the two HST WFC3/IR band filters we used galaxies in the redshift ranges $0.5 < z < 0.85$ and $0.85 < z < 1.27$ for the J and H bands of WFC3/HST, respectively (due to morphological k-correction).

II. We used the ΔBIC to determine the most probable model for every galaxy, splitting the sample into two broad groups; one- and two- component galaxies. The $\Delta\text{BIC} > 25$ cut for the SDSS sample was used to obtain the most probable two-component systems and for the sake of comparison, same was used for the COSMOS sample. We find that the fraction of two-

component galaxies has increased since $z \sim 1$ with the higher stellar mass galaxies being more populated with these systems than the lower mass galaxies. Comparing to the studies from Margalef-Bentabol et al. (2016), Fig.5.5 illustrates an evolution with redshift, such that massive two-component systems increase by a factor ~ 2 from $z \sim 1.5$ to $z \sim 0$ compared to their one-component counterparts. The factor of how the lower stellar mass and higher mass populations differ is ~ 2 from $z \sim 1$ to $z \sim 0.7$. However, we can not rule out that this decrease with redshift could be due lower resolution effects when trying to observe bulges at high redshifts.

III. We utilised the results from the 1DHBM to compare with previous studies as well as to observe any trends within the structural parameters of the galaxies. For the one-component galaxies, for both the stellar mass bins,we find that the galaxies were roughly of equal size at $z \sim 1$ with the higher mass population increasing in size more rapidly from $z \sim 1$ until $z \sim 0$ (Fig.5.8). For one-components progenitors in the lower stellar mass bin, the effective radius has had little growth since $z \sim 1$ and comparing to the results from van Dokkum et al. (2013) this seems to have been the case since $z \sim 2.5$. We also observe that the Sérsic index has also changed since $z \sim 1$. The more massive galaxies always have higher values, but the increase from $z \sim 1$ is the same for both population (factor ~ 2).

IV. Using the results from the 2DHBM we have investigated how the joint probability distributions for the structural parameters has changed since $z \sim 1$. We find that the correlations observed between the Sérsic parameters in the nearby Universe for the one-component galaxies, namely the $R_e - n$ relation (Fig.5.11), reverse at higher redshifts. We show that one-component galaxies at $z \sim 1$ have a negative correlation between R_e and n, with higher n and larger R_e galaxies forming between $z \sim 0.7$ and now.

V. We have studied the structural parameters for the progenitors of two-component galaxies since $z \sim 1$. Using the 1DHBM we find that the effective radii for both the lower and higher mass two-component galaxies remained constant from $z \sim 1$ to $z \sim 0.7$ (see Fig. 5.14). However, between $z \sim 0.7$ and $z \sim 0$, the two mass samples take different evolutionary paths with the higher mass sample increasing in size (by a factor ~ 1.3) and the lower mass sample decreasing by a factor of ~ 0.6 in R_e. The Sérsic index for both mass populations follows a similar trend to that of the one-component galaxies with an increase over the same redshift range. Although, the increase in n is not as steep. One consistency between the samples is the scaling due to stellar mass i.e. the more massive galaxies have larger n values over the entire

redshift range.

VI. The $R_e - n$ plane (Fig.5.16) for the two-component galaxies show similar evolutionary trends to the one-component galaxies at high z ($z \sim 1$) and over time they move onto the local galaxy scaling relation. For the high mass galaxies there is evidence that high n, high R_e bulges are present at $z \sim 1$, with this population growing over the next seven billion years. The lower mass sample follow a similar trend but in the local Universe bulges with $n < 2$ have a higher chance of being more concentrated with lower R_e values.

VII. We take the 1DHBM distribution for the B/T of our two-component galaxies to understand the build-up of the morphologies in the local Universe (Fig.5.18). We find similarities with other local Universe studies with lower mass systems being dominated by $B/T \leq 0.3$ structures at $z \sim 0$. For the high mass sample, bulge dominated galaxies ($B/T \geq 0.7$) do increase since $z \sim 1$, which is in agreement with the semi-empirical simulations from Avila-Reese et al. (2014). However, we do find large discrepancies when comparing our results to that of other high redshift studies (*e.g.,* Bruce et al. 2012), but this might be due to the differing techniques used to separate bulges and discs.

VIII. The disc components in our two-component galaxies appear to have had little change since $z \sim 1$ (Fig.5.21 and 5.22). We observe a subtle growth disc scale lengths for the high mass progenitors. The overall scaling relation between μ_0 and h has remained fixed since $z \sim 1$.

IX. We find little evidence to support that the bulge component is shaping the disc in anyway (Fig.5.23) or vice versa. There is a positive correlation between R_e and h although the probability distribution is approaching Gaussian in shape (i.e., a weak correlation). Since $z \sim 1$, high mass galaxies with $B/T > 0.7$ increase in probability while intermediate B/T galaxies decrease (see Fig.5.24). This is also observed in the lower mass sample,although, the $B/T > 0.7$ galaxies show insignificant evolution.

5.7.1 Evolution of one-component systems

The merging of two roughly equal sized galaxies have long been considered the main channel for the formation of modern elliptical galaxies (Toomre 1977). Simulations of merging disc galaxies show that the disruption and violent relaxation of the stellar component results in the formation of a spheroid (Lynden-Bell 1967; Barnes 1988; Hernquist 1992). Hopkins et al. (2008) used dissipational mergers of disc galaxies to show that the remnant stars were redistributed into a spherical

component after the merging event. However, there has also been counter arguments to this proposing that the merging of two disc galaxies can not form massive elliptical galaxies (Naab & Ostriker 2009).There have been many studies showing that massive galaxies with low star formation rates were compact (i.e., small size with high mass) at $z \geq 2$ (Daddi et al. 2005; Trujillo et al. 2006; van Dokkum et al. 2008; Damjanov et al. 2011; Conselice 2014; van Dokkum et al. 2015). Quiescent galaxies at a fixed stellar mass have been shown to be a factor of ~ 4 smaller at $z = 2$ than at $z = 0$ (van der Wel et al. 2014). Studies have also shown that as the stellar mass evolves, the size growth of a galaxy is even larger (van Dokkum et al. 2010; Patel et al. 2013). In van Dokkum et al. (2015) they studied the formation of massive ($M_* \sim 10^{11} M_\odot$) compact ($R_e \sim$ 1kpc) quiescent galaxies in the redshift range $1.5 \leq z \leq 3$. We find that the size evolution for high mass one-component galaxies (shown in Figure 5.8) are consistent with these previous measurements (e.g. van Dokkum et al. 2008; Williams et al. 2010; Patel et al. 2013; van der Wel et al. 2014).

It has been suggested that the strong evolution of the size of spheroid-dominated galaxies may be driven largely by dissipationless minor mergers (Bournaud et al. 2007; Naab et al. 2007, 2009; Bezanson et al. 2009). Using a cosmological hydrodynamical simulation, Naab et al. (2009) suggested that the galaxies grew by a factor of $\sim$ 1.7 between $z \sim 1$ and $z \sim 0$, which is in good agreement with our finding for the high mass one-component galaxies. However, the overall sizes of their simulated galaxies were a factor ~ 2 lower than what we found. With more recent simulations, Oser et al. (2012) again analysed hydrodynamic cosmological simulations of the present day massive ($\log(M_*/M_\odot > 10.8)$) galaxies. It was suggested that their progenitors were very compact (projected half-mass radii of $\leq$ 1.3kpc) and that they grew in size by a factor of ~ 2.5 since $z \sim 1$. This growth factor is larger than our measurements but is consistently within the probability distributions. Figure 5.11 shows at $z \sim 1$ there was a high probability that the high mass sample were compact with high n (> 2) values. By $z \sim 0$ the high mass sample appear to have more extended radii as well as $n \sim 4$ with no evidence for the existence of the compact $z \sim 1$ population. This is also in agreement with previous studies (e.g. Trujillo et al. 2009; Taylor et al. 2009). According to Oser et al. (2012), the subsequent evolution of the galaxies since $z \leq 2$ is dominated by accretion of stars in satellite stellar systems. They also show that the accreted stellar systems settle into the outer parts of the galaxies, resulting in an increase of their size. However, some studies have suggested that the minor merger rate may not be sufficient to account for the observed evolution in the size-mass relation (Newman et al. 2012; Shankar et al. 2013).

Porter et al. (2014) implemented a recipe for computing spheroid sizes and velocity dispersions

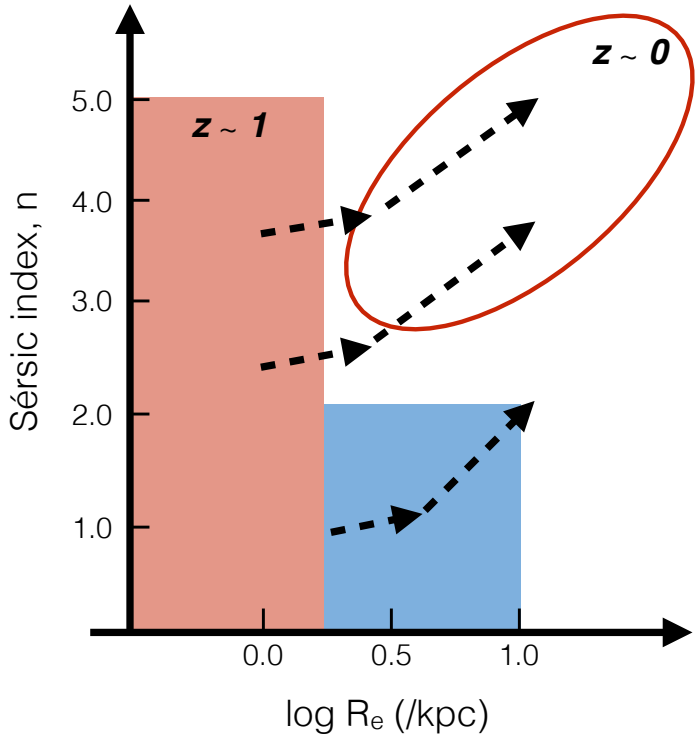


Figure 5.26: A schematic of the evolution for the one-component galaxies since $z \sim 1$. The arrows show the overall growth of the galaxies shown by the results earlier in this chapter. The compact progenitors (red box) may take one of two paths: (1) either they grow rapidly due to dry minor mergers which deposit material onto their outer parts (leading to the local high mass one-component scaling relation shown by the red ellipse) or (2) infalling gas forms a disc around the spheroid becoming $z \sim 0$ bulge-dominated two-component galaxies. The high mass, extended, one-component galaxies at $z \sim 1$ could become the discs of the modern disc-dominated galaxies and most likely form a bulge (blue box). Whereas evidence suggests that the lower mass extended galaxies at $z \sim 1$ could become the population of modern pure disc galaxies. However, this may be dependent of the enviroment around the galaxy.

based on a simple analytic model using Semi-Analytic Modelling (SAM) including the effects of gas dissipation. With a model in which spheroids grow only via mergers, they find that there is difficulty in the recovery of the observed number density of intermediate-mass spheroid-dominated galaxies at $z \sim 0$. They also show that the model predicts rapid size evolution of spheroid-dominated galaxies since $z \sim 2$. This is in agreement with our high mass progenitors as we also see a rapid size evolution (Fig.5.8). In this picture, dissipation plays a major role in explaining the different slopes and evolutions of the size-mass relations for the spheroid galaxies in relation to the discs. Lower mass spheroids have lower mass progenitors, which have higher gas fractions at all redshifts. More gas means more dissipation and smaller remnants, thus a steeper size-mass relation. Figure 5.25 is in agreement with this picture where the lower mass one-component progenitors will follow a similar size growth to the high mass galaxies but are smaller at all redshifts. To summarise, it seems apparent with the recent semi-analytic models (Porter et al. 2014) and cosmological simulations (Oser et al. 2012) that both dissipational collapse and mergers are likely involved in the formation of ellipticals in the near-by Universe.

If the compact one-component galaxies become modern day ellipticals, can the more extended one-components at $z \sim 1$ become the discs we observe in the local Universe? Somerville et al. (2008) shown that their CDM—based model of disc formation produces a weak redshift evolution of the disc size out to $z \sim 1$ which is in good agreement with our observations (see Fig. 5.21). This result is also in agreement with the past study on the COSMOS survey by Sargent et al. (2007). They found that the number density of discs with half-light radii between 5 and 7 kpc is almost constant since $z \sim 1$. This portrays a picture where the discs have a fixed relationship between their stellar mass and their radial size from $z \sim 1$ to $z \sim 0$.

The more gradual disc evolution in Somerville et al. (2008) results primarily from the use of realistic NFW halo profiles and from the incorporation of the redshift evolution of the halo concentration-mass relation predicted by N-body simulations. They demonstrated that less concentrated halos produce larger discs because of the lower mass and weaker gravitational forces in the central parts of the halo. Thus, the collapse of baryons produces less concentrated dark matter profiles, and a more extended disc. The trend toward lower concentrations at earlier times therefore counteracts the decreasing virial radii. Out to about $z \sim 1$, these competing effects nearly cancel out, leading to the weak evolution seen between the size and mass of disc galaxies. In our study we can see that the high mass, extended galaxies with low n values have diminished since $z \sim 1$. This could be a sign that they become the discs of the two-component galaxies at $z \sim 0$(see Fig. 5.25). Whereas,

lower mass extended one-components appear to remain over the redshift range with little change in both their size and shape thus suggesting that they become the pure disc galaxies we observe in the local Universe.

5.7.2 Formation of modern day bulges

Recent studies have also revealed that bulges (or the progenitor of bulges) are present in the Universe by $z \sim 1$ (Huetas-Company et al. 2015; Margalef-Bentabol et al. 2016; Sachdeva et al. 2017; Tadaki et al. 2017). However, the formation and evolution of these components are thought to be very complex with different processes likely to be at work. In this final section we put our results for the bulges of two-component galaxies into the frame work of previous observational and theoretical studies. There are currently three leading theories of bulge formation: i) the accretion of gas onto a dense spheroid after a major merger, ii) the building of a bulge through the accretion of smaller satellites, i.e., minor mergers, and iii) through various internal secular processes such as disc instabilities and the formation of a bar component (see Fig. 5.27).

Massive bulges, which are predominately found in S0 and Sa galaxies, are known to share many properties with ellipticals in the local Universe (for a review see Wyse et al. 1997). In relation to their structures both ellipticals and massive bulges closely follow a Sérsic profile with high concentrations of light (i.e., roughly $n \geqslant 3.5$). In addition to this, massive bulges and ellipticals obey similar colour-magnitude relations (Balcells & Peletier, 1994), similar metallicity-luminosity relations (Jablonka et al. 1996), similar fundamental plane relations (Bender et al. 1992), and both systems follow a close relation between the mass of the super massive black hole in their centres and the velocity dispersions (see Tremaine et al. 2002 and references therein). These observations suggest both systems have followed a similar formation path. It has long been theorised that a galaxy will either become a bulge (and disc) or an elliptical based on whether the disc can survive a merging event or that a new disc can be formed by the accretion of gas (e.g. Kauffmann et al. 1993; Baugh et al. 1996; Governato et al. 2009).

In the more recent semi-empirical cosmological simulations of Zavala et al. (2012) and Avila-Reese et al. (2014), they show the growth of bulges with realistic merger rates as a function of time while the galaxy is in a growing a CDM halo. They suggested that this merger building scenario can predict the classical- and the disc-like bulges alike. Their model includes three ways a bulge can they acquire their stars: (1) from the merged secondary, (2) from the primary disc due to instabilities induced by the mergers, and (3) through stars formed *in situ* from the gas that is funnelled to the

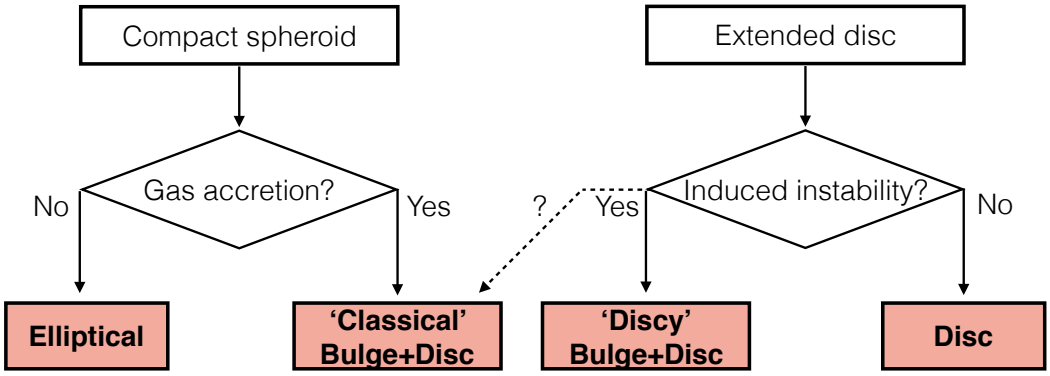


Figure 5.27: A logic flow for a simplified description of how to form the broad classification in the local Universe.

centre during merging. In Figure 5.18 we compare to the predictions made by Avila-Reese et al. (2014) for galaxies with $M_* > 10^{11} M_\odot$. The clear discrepancy is with lower B/T galaxies, however, intermediate and high B/T galaxies follow similar trends to their predictions. Our $B/T > 0.7$ high mass galaxies are consistently lower by a factor of ~ 1.3 from $z \sim 1$ with the intermediate galaxies ($0.3 < B/T < 0.7$) falling lower by a factor ~ 2.5. A comparison of B/T ratios must be done carefully, though. A number of works have noted that the B/T determined through kinematic decomposition are systematically larger than those determined through photometric decompositions (Governato et al. 2009; Scannapieco et al. 2010; Marinacci et al. 2013; Aumer et al. 2014; Christensen et al. 2014). Thus if we take this into consideration our results for galaxies with $B/T > 0.3$ begin to be in agreement. Numerical simulations also show that when the mass of the satellite is of the order of 10% of the disc, it can sink to the centre of the potential well in a few Gyr (e.g. Walker et al. 1996). Aguerri et al. (2001) have shown that the accretion of multiple satellite galaxies may create a relation between the bulge mass and n. They also suggest that after the accretion of a satellite, the bulge grows and n increases proportional to the satellite mass. This is in agreement with Figure 5.14 where we see the high mass progenitors increasing their size and n values gradually over time.

The secular evolution of a galaxy may also give rise to the bulge population at $z \sim 0$ (e.g., Kormendy & Kennicutt, 2004; Eliche-Moral et al. 2006; López-Sanjuan et al. 2009; Hopkins et al. 2010; Bournaud et al. 2014; Kaviraj 2014). It is often assumed that smaller bulges located in disc dominated spiral systems formed in a different fashion to their larger more compact counterparts (see Costantin et al. 2017). This has been motivated by the fact that they share similar properties to their host discs such as how their surface brightness are close to exponentials (Kormendy

& Kennicutt, 2004). Observations have shown that there is a correlation between R_e and h in the local Universe adding support bulge growth through secular evolution and the coevolution of bulges and discs (Courteau et al. 1996; Aguerri et al. 2005; Carollo et al. 2007). The results presented here (Figure 5.23) show that there is indeed a subtle correlation between the bulge and disc size since $z \sim 1$. We can also see that bulges and disc in higher mass galaxies are situated at the higher end at $z \sim 0$ but share similar sizes to the lower mass sample earlier in the past. Instability-driven bulge growth through migration of star-forming clumps have been shown to live long enough to reach the centres of high redshift galaxies to eventually form discy-bulges (Bournaud et al. 2014). It was shown in Bournaud et al. (2014) that these clumps could grow a bulge to a similar bulge mass seen in present-day galaxies. Our analysis suggests that a joint role between all the discussed evolutionary scenarios is probably the likely mechanisms forming bulges in the local Universe and further studies with deep imaging (*JWST*) would be required.

6

Conclusions

In the is thesis, we have explored the structural parameters of a sample galaxies to better understand their evolution from $z \sim 1$ until now. The approach taken has been largely observational with the use of the new 2D photometric decomposition code PHI with the application of a Hierarchical Bayesian Model (HBM) to better capture the underlying population trends of the galaxy sample. The HBM has given us the opportunity to look at structure in the Universe in a new light, as for the first time we are able to place proper probabilistic constraints on the population parameters. We now give a summary for each chapter in this thesis and end with possible extensions to the work.

6.1 Thesis summary

Chapter 2:

I. In this chapter 2 we introduced a new fitting algorithm (PHI) to perform 2D photometric decompositions of galaxy images from a Bayesian perspective. The new method offers a number of significant advantages for estimating surface brightness profile parameters. The new algorithm addresses the five common reasons for previous minimisation methods to fail e.g. i)

Local minima trapping, ii) unrealistic solutions, iii) reversal of components, iv) indecisiveness to which model to use , and v) bad representation of the final errors.

II. PHI incorporates a triple layer approach. The first layer uses a blocked adaptive Metropolis algorithm to obtain an estimate of the scale for each parameter in the chain. The second layer uses an adaptive Metropolis algorithm with the purpose of estimating the target covariance matrix. We assume the proposed distribution can be described as multivariate normal distribution. The final level uses this calculated covariance matrix to quickly and effectively explore the parameter space reducing the chances of a local minima trap.

III. We implemented a number of priors that aim to allow the parameters to stay realistic and physical (i.e. positive in the case for the dimensions and intensities). These priors are better understood as boundary regions similar to the a filtering process used in past work to remove non-physical parameter outcomes.

IV. We have also outlined our model comparison methodology that helps us determine the most probable morphology (or model) for a galaxy. This involves using Bayesian residual plots for visual inspection used in conjunction with a quantitative interpretation, the Bayesian Information Criteria.

Chapter 3:

I. We have described a synthetic galaxy imager algorithm. We implement a range of cosmological laws within the code. To provide a more realistic image, synthetic stellar populations have also been included in the image generation. Four different star formation histories were used to calculate the apparent magnitude in specific bands. The imager also uses a range of surface brightness profiles to estimate the shape of the intensity received from an emitting source. Furthermore, the images simulates realistic observations by convolving the image with a Point Spread Function (PSF) as well as simulating the background noise levels.

II. We generate a sample of synthetic galaxies to investigate the effects of observational attributes along with any internal code biases that could effect the final output. We find that the main systematic effecting the methodology is the relative size of the effective radius compared to the PSF size. Also the limiting resolution for smaller inner components drives the majority of the uncertainties.

III. We have also applied PHI to a sample of SDSS galaxies. The aim of this was to determine how consistent the new algorithm was compared to previous algorithms in obtaining parameter estimates. The first step showed that under the same image conditions i.e. the same galaxies with the same systematics, PHI and GASP2D achieved consistent results with minor scatter. This validates both algorithms when analysing galaxy structures in the nearby universe. We next compared the results of PHI with the results from Gadotti (2009). We show that parameter values differed significantly when comparing to images of the same galaxy but with different PSFs and different data releases.

Chapter 4:

I. In this chapter, we show how Hierarchical Bayesian models (HBM) can be used to help estimate global structural properties of the sample. The iterative methodology of population modelling allows us to reduce the uncertainties due to the so-called *shrinkage*.

II. We analysed the structural components of a sample of 454 SDSS galaxies to investigate possible scaling relations. The galaxy sample was photometrically decomposed using one and two-component galaxy models. Using this information, together with the HBM, we constrain the mechanisms of bulge and disc assembly in the nearby Universe.

III. We find that the Sérsic parameters of both the one and two-component galaxies follow similar trends despite their location on the Kormendy relation or their estimated n value. This suggests that fundamentally the formation of bulges is driven by their relative size and possibly mass.

IV. Correlations between the Sérsic and exponential parameters in our two-component galaxies were found but only between R_e and h, meaning that larger bulges are more likely to reside in large discs. Although, stronger correlations are present when we look at the ratio between R_e and h verses μ_e and n. We show that more disc dominated galaxies have a higher probability of being brighter and having lower n values.

V. The interpretation of our results can favour scenarios of both hierarchical clustering and that of secular evolution. Thus an investigation of these relations between higher redshifts might help to understand their formation scenarios.

VI. We have also looked into how the classification of the sample into one and two-component galaxies may change the results. We compare use of the ΔBIC (> 25) threshold, to the results from machine learning which separates the galaxies into ellipticals and discs. We find that

when we separate them according to the ΔBIC we include pure disc galaxies in the one component galaxy sample. These are removed when using the machine-learning approach due to the inclusion of other galaxy properties such as the colour.

Chapter 5:

I. In this chapter 5 we studied the structural evolution of 1368 COSMOS/3DHST galaxies in the redshift range $0.5 \leq z \leq 1.27$ and compared them to a local sample of 454 SDSS galaxies. We have applied a HBM including the posterior probabilities from 2D photometric bulge-disc decompositions to achieve a full probabilistic representation of galaxy structures. The non-parametric nature of our HBM means we can re-populate the parameter space based on the population probability function. We have then used this to compare directly with previous observational and theoretical studies.

II. We used stellar population models to obtain an estimate of the mass progenitors for galaxies with masses greater than that of the Milky-Way ($M_* > 10^{10.66} M_\odot$) in the local universe. The same was done for nearby galaxies with masses $10^{10} M_\odot < M_* \leq 10^{10.66} M_\odot$. As our local sample was analysed in the i-band of SDSS, we had to take morphological k-correction into account meaning that we had a limited redshift window to study galaxies in the COSMOS field in. Thus, to properly compare our local sample ($0.04 < z < 0.06$) in the SDSS i-band to the two HST WFC3/IR band filters we used images in the redshift ranges $0.5 < z < 0.85$ and $0.85 < z < 1.27$ for the J and H bands of WFC3/HST, respectively.

III. We used the ΔBIC to determine the most probable model for an object, splitting the sample into two broad groups; one- and two- component galaxies. ΔBIC > 25 was used as a threshold to obtain the most probable two-component systems. We find that the fraction of 2-component galaxies has increased since $z \sim 1$ with the higher stellar mass galaxies being more populated with these systems than the lower mass galaxies. Comparing to the studies from Margalef-Bentabol et al. (2016), Fig.5.5 illustrates an evolution with redshift, such that massive two-component systems increase by a factor ~ 2 from $z \sim 1.5$ to $z \sim 0$ compared to their one-component counterparts. The factor of how the lower stellar mass and higher mass populations differ is ~ 2 from $z \sim 1$ to 0.7.

IV. We utilised the results from the one-dimensional HBM (1DHBM) to better compare to previous studies as well as to observe any trends within the data. For the one-component galaxies, we find the galaxies were roughly of equal size at $z \sim 1$ with the higher mass population increasing

in size more rapidly between $z \sim 1$ and $z \sim 0$ (Fig.5.8). For the lower mass one-component progenitors the effective radius has had little growth since $z \sim 1$ and comparing to the results from van Dokkum et al. (2013) this seems to have been the case since $z \sim 2.5$. We also observe that the Sérsic index has also changed since $z \sim 1$, with the more massive galaxies always having higher values but the factor of ~ 2 increase is observed for both samples.

V. Using the results from the two-dimensional HBM (2DHBM) we investigated how the joint probability distributions for the structural parameters has changed since $z \sim 1$. We find that the correlations observed between the Sérsic parameters in the nearby Universe for the one-component galaxies, namely the $R_e - n$ relation (Fig.5.11), reverse at higher redshifts. We show that one-component galaxies at $z \sim 1$ have a negative correlation between R_e and n with higher n, larger R_e galaxies forming between $z \sim 0.7$ and now.

VI. Using the same methods, we have studied the structural parameters for the progenitors of two-component galaxies. Using the 1DHBM we find that the effective radii for both the lower and higher mass two-component galaxies remained constant from $z \sim 1$ to $z \sim 0.7$ (see Fig. 5.14). However, between $z \sim 0.7$ and $z \sim 0$, the two populations take differing paths with the higher mass sample increasing in size (by a factor ~ 1.3) and the lower mass sample decreasing by a factor of ~ 0.6 in R_e. The Sérsic index for both mass populations follows a similar trend to that of the one-components in that there is an increase over the same redshift range. The increase in n is not as steep, however, the more massive galaxies have larger n values over the entire redshift range.

VII. The $R_e - n$ plane (Fig.5.16) for the two-component galaxies present similar evolutionary trends to that of the one-components in that objects at high z ($z \sim 1$) have lower n and higher R_e values. For the more massive galaxies there is evidence that high n, high R_e bulges are present at $z \sim 1$, with this population growing over the next seven billion years. The lower mass sample follow a similar trend but in the local Universe bulges with $n < 2$ have a higher chance of being more concentrated with lower R_e values.

VIII. We take the 1DHBM distribution for the B/T of our two-component galaxies to understand the build-up of morphologies in the local Universe (Fig.5.18). We find similarities with other local Universe studies with lower mass systems mostly being disc dominated structures ($B/T \leq 0.3$). Since $z \sim 1$ these $B/T \leq 0.3$ objects are the most likely for both the massive and lower mass galaxies. Bulge dominance ($B/T \geq 0.7$) does increase though for the massive galaxies since

$z \sim 1$, which is in agreement with the semi-empirical simulations from Avila-Reese et al. (2014).

IX. The disc relations in our two-component galaxies are observed to remain constant over the redshift range $0 < z < 1$ (Fig.5.21 and 5.22). We have observed a subtle growth in the disc scale lengths for the massive and lower mass progenitors.

X. We find little evidence to support that bulges are shaping their discs in anyway (Fig.5.23) or vice versa. There is a positive correlation between R_e and h, although, the probability distribution is approaching Gaussian in shape (i.e. a weak correlation). Dominate bulges ($B/T > 0.7$) appear more prevalent in the massive galaxies around $z \sim 0.7$ (see Fig.5.24). For the lower mass sample between $z \sim 0.7$ and $z \sim 0$ the $B/T > 0.7$ galaxies evolve insignificantly, which is in contrast to the massive galaxy population.

6.2 Outlook

In the future, to aid the structural analysis of highly detailed galaxy images from new telescopes such as LSST and JWST, a fully probabilistic methodology will need to be applied. The Bayesian MCMC algorithm that we have developed will tailor to this need. With the advent release of the algorithm and the updates to follow including new component profiles (e.g. for galaxy bars) as well as the potential for to include new complex hierarchical models, it is clear that the methodology will remain relevant and of high impact for many years to come.

The continued study of high redshift galaxies will yield more clues about the limits of galaxy structures, expanding our picture the mechanisms at play in the Universe. In particular, as touched upon in Chapter 5, the formation of bulges is only really just beginning as telescopic resolution increases. In this thesis, we have glimpsed the build-up of bulges but only with a multi-observational platform, linking photometric and dynamical measurements, can we really begin to fully understand the formation process.

There is also evidence for morphology driven quenching. Utilising the information given by the 3D-HST survey we can begin to investigate whether bulges with higher Sérsic indexes are more likely to be quenched over cosmic time. Linking the structural parameters found in this thesis with the physical parameters of their host galaxies we can begin to answer some of the more fundamental questions remaining.

Bibliography

Andredakis, Y. C., Peletier, R. F., & Balcells, M. 1995, MNRAS, 275, 874

Athanassoula, E. 1992, MNRAS, 259, 345

——. 2005, MNRAS, 358, 1477, astro-ph/0502316

——. 2008, MNRAS, 390, L69, 0808.0016

Balcells, M., Graham, A. W., Domínguez-Palmero, L., & Peletier, R. F. 2003, ApJ, 582, L79, astro-ph/0212184

Baugh, C. M., Cole, S., & Frenk, C. S. 1996, MNRAS, 283, 1361, astro-ph/9602085

Begeman, K. G. 1989, A& A, 223, 47

Benson, A. J. 2010, 'Physics Reports', 495, 33, 1006.5394

Bernardi, M. et al. 2003, AJ, 125, 1849, astro-ph/0301624

Bertin, G., & Lin, C. C. 1996, Spiral structure in galaxies a density wave theory

Bica, E. 1988, A&A, 195, 76

Blumenthal, G. R., Faber, S. M., Primack, J. R., & Rees, M. J. 1984, Nature, 311, 517

Bottema, R. 1989, A&A, 221, 236

——. 1993, A&A, 275, 16

Bruce, V. A. et al. 2012, MNRAS, 427, 1666, 1206.4322

——. 2014, ArXiv e-prints, 1405.1736

Bruzual, G., & Charlot, S. 2003, MNRAS, 344, 1000, astro-ph/0309134

Buitrago, F., Trujillo, I., Conselice, C. J., & Häußler, B. 2013, MNRAS, 428, 1460, 1111.6993

Caon, N., Capaccioli, M., & D'Onofrio, M. 1993, MNRAS, 265, 1013, astro-ph/9309013

Capaccioli, M., & Caon, N. 1989, in European Southern Observatory Conference and Workshop Proceedings, Vol. 31, ESO/ST-ECF Data Analysis Workshop, ed. P. J. Grosbøl, F. Murtagh, & R. H. Warmels, 107–126

Carollo, C. M., Ferguson, H. C., & Wyse, R. F. G., eds. 1999, The formation of galactic bulges

Carollo, C. M., Stiavelli, M., Seigar, M., de Zeeuw, P. T., & Dejonghe, H. 2002, AJ, 123, 159, astro-ph/0110281

Ciotti, L. 1991, AAP, 249, 99

Cole, S., Lacey, C. G., Baugh, C. M., & Frenk, C. S. 2000, MNRAS, 319, 168, astro-ph/0007281

Combes, F. 2007, in IAU Symposium, Vol. 235, IAU Symposium, ed. F. Combes & J. Palouš, 19–23, astro-ph/0608612

Costantin, L., Méndez-Abreu, J., Corsini, E. M., Morelli, L., Aguerri, J. A. L., Dalla Bontà, E., & Pizzella, A. 2017, A & A, 601, A84, 1703.02976

de Jong, R. S. 1996, A& A, 313, 377, astro-ph/9604010

de Souza, R. E., Gadotti, D. A., & dos Anjos, S. 2004, ApJ, 153, 411, astro-ph/0404103

de Vaucouleurs, G. 1948, Annales d'Astrophysique, 11, 247

——. 1956, Survey of bright galaxies south of -35 deg. declination with the 30-inch Reynolds reflector (1952-1955) (Pub)

Debattista, V. P. 2006, in Astronomical Society of the Pacific Conference Series, Vol. 352, New Horizons in Astronomy: Frank N. Bash Symposium, ed. S. J. Kannappan, S. Redfield, J. E. Kessler-Silacci, M. Landriau, & N. Drory, 161, astro-ph/0601277

Debattista, V. P., Carollo, C. M., Mayer, L., & Moore, B. 2005, ApJ, 628, 678, astro-ph/0504530

Dieleman, S., Willett, K. W., & Dambre, J. 2015, MNRAS, 450, 1441, 1503.07077

Eke, V. R., Cole, S., & Frenk, C. S. 1996, MNRAS, 282, 263, astro-ph/9601088

Ellis, R. S., Abraham, R. G., & Dickinson, M. 2001, ApJ, 551, 111, astro-ph/0010401

Emsellem, E. 1998, in IAU Symposium, Vol. 184, The Central Regions of the Galaxy and Galaxies, ed. Y. Sofue, 397

Erwin, P. 2004, A&A, 415, 941, astro-ph/0310806

Erwin, P., Beckman, J. E., & Pohlen, M. 2005, ApJ, 626, L81, astro-ph/0505216

Falcón-Barroso, J. et al. 2006, MNRAS, 369, 529, astro-ph/0603161

Falcón-Barroso, J., Peletier, R. F., & Balcells, M. 2002, MNRAS, 335, 741, astro-ph/0205176

Fall, S. M., & Efstathiou, G. 1980, MNRAS, 193, 189

Fathi, K. 2010, ApJ, 722, L120, 1009.2692

Fathi, K., Allen, M., Boch, T., Hatziminaoglou, E., & Peletier, R. F. 2010, MNRAS, 406, 1595, 1004.1507

Fathi, K., & Peletier, R. F. 2003, A& A, 407, 61, astro-ph/0305326

Fisher, D. B., & Drory, N. 2010, ApJ, 716, 942, 1004.5393

Freeman, K. C. 1970, ApJ, 160, 811

Gadotti, D. A. 2008, in IAU Symposium, Vol. 245, IAU Symposium, ed. M. Bureau, E. Athanassoula, & B. Barbuy, 117–120, 0708.2842

Gadotti, D. A. 2009, MNRAS, 393, 1531, 0810.1953

Geweke, J. 1992, in IN BAYESIAN STATISTICS (University Press), 169–193

Giordani, P., & Kohn, R. 2010, Journal of Computational and Graphical Statistics, 19, 243, http://dx.doi.org/10.1198/jcgs.2009.07174

Gnedin, N. Y., & Hui, L. 1998, MNRAS, 296, 44, astro-ph/9706219

Gorgas, J., Jablonka, P., & Goudfrooij, P. 2007, A&A, 474, 1081, 0707.0407

Graham, A. W. 2001, AJ, 121, 820, astro-ph/0011256

Graham, A. W., & Driver, S. P. 2005, PSAA, 22, 118, astro-ph/0503176

Gunn, J. E. 1982, in Astrophysical Cosmology Proceedings, ed. H. A. Brueck, G. V. Coyne, & M. S. Longair, 233–259

Haario, H., Saksman, E., & Tamminen, J. 2001, Bernoulli, 7, 223

Hammer, F., Gruel, N., Thuan, T. X., Flores, H., & Infante, L. 2001, ApJ, 550, 570, astro-ph/0011218

Hasan, H., Pfenniger, D., & Norman, C. 1993, ApJ, 409, 91

Heller, C. H., & Shlosman, I. 1994, ApJ, 424, 84

Hogg, D. W. 1999, ArXiv Astrophysics e-prints, astro-ph/9905116

Hogg, D. W., Baldry, I. K., Blanton, M. R., & Eisenstein, D. J. 2002, ArXiv Astrophysics e-prints, astro-ph/0210394

Hubble, E. P. 1926, AJ, 64, 321

Huertas-Company, M., Aguerri, J. A. L., Tresse, L., Bolzonella, M., Koekemoer, A. M., & Maier, C. 2010, A&A, 515, A3, 1002.3076

Huertas-Company, M. et al. 2015, ApJS, 221, 8, 1509.05429

Huertas-Company, M., Rouan, D., Tasca, L., Soucail, G., & Le Fèvre, O. 2008, A&A, 478, 971, 0709.1359

Humason, M. L., Mayall, N. U., & Sandage, A. R. 1956, AJ, 61, 97

Idiart, T. P., de Freitas Pacheco, J. A., & Costa, R. D. D. 1996, AJ, 112, 2541

Jablonka, P., Bica, E., Bonatto, C., Bridges, T. J., Langlois, M., & Carter, D. 1998, A&A, 335, 867, astro-ph/9802272

Jablonka, P., Bica, E., Pelat, D., & Alloin, D. 1996, A&A, 307, 385

Jablonka, P., Gorgas, J., & Goudfrooij, P. 2007, A&A, 474, 763, 0707.0561

Jiang, F., & van den Bosch, F. C. 2014, MNRAS, 440, 193, 1311.5225

Kauffmann, G. 1996, MNRAS, 281, 475, astro-ph/9512123

Kent, S. M. 1985, ApJ, 59, 115

Khosroshahi, H. G., Wadadekar, Y., & Kembhavi, A. 2000, ApJ, 533, 162, astro-ph/9911402

Kim, T., Gadotti, D. A., Athanassoula, E., Bosma, A., Sheth, K., & Lee, M. G. 2016, MNRAS, 462, 3430, 1607.08245

Komatsu, E. et al. 2009, Astrophysical Journal series, 180, 330, 0803.0547

Koo, D. C., Datta, S., Willmer, C. N. A., Simard, L., Tran, K.-V., & Im, M. 2005, ApJ, 634, L5, astro-ph/0510657

Kormendy, J. 1977, ApJ, 217, 406

——. 1982, ApJ, 257, 75

Kormendy, J., & Illingworth, G. 1982, ApJ, 256, 460

Kormendy, J., & Kennicutt, Jr., R. C. 2004, Ann. Re. A & A, 42, 603, astro-ph/0407343

Krist, J., Hook, R., & Stoehr, F. 2010, Tiny Tim: Simulated Hubble Space Telescope PSFs, 1010.057

Kuhlen, M., Diemand, J., Madau, P., & Zemp, M. 2008, Journal of Physics Conference Series, 125, 012008, 0810.3614

La Barbera, F., Busarello, G., Merluzzi, P., Massarotti, M., & Capaccioli, M. 2003, ApJ, 595, 127, astro-ph/0212294

Liddle, A. R. 1999, in High Energy Physics and Cosmology, 1998 Summer School, ed. A. Masiero, G. Senjanovic, & A. Smirnov, 260, astro-ph/9901124

Lin, C. C., & Shu, F. H. 1964, ApJ, 140, 646

Lintott, C. et al. 2011, MNRAS, 410, 166, 1007.3265

Lintott, C. J. et al. 2008, MNRAS, 389, 1179, 0804.4483

Lisker, T. 2008, ApJS, 179, 319, 0807.1531

MacArthur, L. A., Ellis, R. S., Treu, T., U, V., Bundy, K., & Moran, S. 2008, apJ, 680, 70, 0711.0238

Martinez-Valpuesta, I., Shlosman, I., & Heller, C. 2006, ApJ, 637, 214, astro-ph/0507219

Masters, K. L. et al. 2010, MNRAS, 405, 783, 0910.4113

McKee, C. F., & Ostriker, E. C. 2007, ARAA, 45, 565, 0707.3514

Mehlert, D., Thomas, D., Saglia, R. P., Bender, R., & Wegner, G. 2003, A&A, 407, 423, astro-ph/0306219

Méndez-Abreu, J., Aguerri, J. A. L., Corsini, E. M., & Simonneau, E. 2008a, A&A, 487, 555, 0710.5466

Méndez-Abreu, J., Corsini, E. M., Debattista, V. P., De Rijcke, S., Aguerri, J. A. L., & Pizzella, A. 2008b, ApJ, 679, L73, 0804.3589

Méndez-Abreu, J., Simonneau, E., Aguerri, J. A. L., & Corsini, E. M. 2010, A&A, 521, A71, 1006.5449

Mo, H., van den Bosch, F. C., & White, S. 2010, Galaxy Formation and Evolution

Morelli, L., Cesetti, M., Corsini, E. M., Pizzella, A., Dalla Bontà, E., Sarzi, M., & Bertola, F. 2010, A&A, 518, A32, 1004.2190

Mortlock, A. et al. 2013, MNRAS, 433, 1185, 1305.2204

Noguchi, M. 1987, Sci. Rep. Tohoku Univ. Eighth Ser., 7, 321

Oke, J. B., & Gunn, J. E. 1983, ApJ, 266, 713

Oke, J. B., & Sandage, A. 1968, ApJ, 154, 21

Peng, C. Y., Ho, L. C., Impey, C. D., & Rix, H.-W. 2002, AJ, 124, 266, astro-ph/0204182

Phillips, A. C., Illingworth, G. D., MacKenty, J. W., & Franx, M. 1996, AJ, 111, 1566

Pignatelli, E., Fasano, G., & Cassata, P. 2006, A&A, 446, 373

Pizzella, A., Corsini, E. M., Morelli, L., Sarzi, M., Scarlata, C., Stiavelli, M., & Bertola, F. 2002, ApJ, 573, 131, astro-ph/0203038

Pohlen, M., Dettmar, R.-J., Lütticke, R., & Aronica, G. 2002, A & A, 392, 807

Pohlen, M., & Trujillo, I. 2006, A & A, 454, 759, astro-ph/0603682

Raine, D., & Thomas, T. 2001, An introduction to the science of cosmology

Renzini, A. 1999, in The Formation of Galactic Bulges, ed. C. M. Carollo, H. C. Ferguson, & R. F. G. Wyse, 9, astro-ph/9902108

Rest, A., van den Bosch, F. C., Jaffe, W., Tran, H., Tsvetanov, Z., Ford, H. C., Davies, J., & Schafer, J. 2001, AJ, 121, 2431, astro-ph/0102286

Roberts, G. O., & Rosenthal, J. S. 2001, Statist. Sci., 16, 351

Roberts, G. O., & Rosenthal, J. S. 2007, J. Appl. Probab., 44, 458

Roberts, G. O., & Rosenthal, J. S. 2009, Journal of Computational and Graphical Statistics, 18, 349

Robotham, A., Taranu, D. S., Tobar, R., Moffett, A., & Driver, S. 2017, MNRAS, 466, 1513, 1611.08586

Roškar, R., Debattista, V. P., Quinn, T. R., Stinson, G. S., & Wadsley, J. 2008a, ApJ, 684, L79, 0808.0206

Roškar, R., Debattista, V. P., Stinson, G. S., Quinn, T. R., Kaufmann, T., & Wadsley, J. 2008b, ApJ, 675, L65, 0710.5523

Sagan, C. 1981, Cosmos.

Sandage, A., & Bedke, J. 1994, The Carnegie Atlas of Galaxies. Volumes I, II.

Schwarz, G. 1978, Ann. Statist., 6, 461

Sersic, J. L. 1968, Atlas de galaxias australes (Pub)

Seth, A. C., Dalcanton, J. J., Hodge, P. W., & Debattista, V. P. 2006, AJ, 132, 2539, astro-ph/0609302

Shaw, D. J., & Mota, D. F. 2008, Astrophysical Journal series, 174, 277, 0708.0868

Shen, J., & Sellwood, J. A. 2004, in Bulletin of the American Astronomical Society, Vol. 36, American Astronomical Society Meeting Abstracts, 122.04

Sheth, K. et al. 2008, ApJ, 675, 1141, 0710.4552

Simard, L. 1998, in Astronomical Society of the Pacific Conference Series, Vol. 145, Astronomical Data Analysis Software and Systems VII, ed. R. Albrecht, R. N. Hook, & H. A. Bushouse, 108

Simard, L. et al. 2002, ApJ, 142, 1, astro-ph/0205025

Simmons, B., Lintott, C., Masters, K., Willett, K., Kartaltepe, J. S., Closson Ferguson, H., Faber, S. M., & Galaxy Zoo Team, C. T. 2016, in American Astronomical Society Meeting Abstracts, Vol. 227, American Astronomical Society Meeting Abstracts, 342.42

Simpson, J. M. et al. 2014, ApJ, 788, 125, 1310.6363

Slyz, A. D., Devriendt, J. E. G., Silk, J., & Burkert, A. 2002, MNRAS, 333, 894, astro-ph/0106058

Sommer-Larsen, J., Götz, M., & Portinari, L. 2003, Apj, 596, 47, astro-ph/0204366

Swaters, R. A., Madore, B. F., & Trewhella, M. 2000, ApJ, 531, L107, astro-ph/0001277

Tamminen, J. 1999, Geophysical Publications, 82

Thomas, D., Maraston, C., Bender, R., & Mendes de Oliveira, C. 2005, ApJ, 621, 673, astro-ph/0410209

Tully, R. B., & Fisher, J. R. 1987, Sky and Telescope, 74, 612

van der Kruit, P. C. 1987, A & A, 173, 59

van der Kruit, P. C., & Freeman, K. C. 2011, ARAA, 49, 301, 1101.1771

van der Kruit, P. C., & Searle, L. 1981a, A & A, 95, 105

——. 1981b, A & A, 95, 116

Vihola, M. 2012, Statistics and Computing, 22, 997

Wild, V., Kauffmann, G., Heckman, T., Charlot, S., Lemson, G., Brinchmann, J., Reichard, T., & Pasquali, A. 2007, MNRAS, 381, 543, 0706.3113

Willett, K. W. et al. 2013, MNRAS, 435, 2835, 1308.3496

Wyse, R. F. G., Gilmore, G., & Franx, M. 1997, Ann.Rev.A & A, 35, 637, astro-ph/9701223

Yoshii, Y., & Sommer-Larsen, J. 1989, MNRAS, 236, 779